



FACULTEIT LETTEREN EN WIJSBEGEERTE

## **VADEMECUM VOOR TAALKUNDIG ONDERZOEK**

Universiteit Gent

Vakgroep Taalkunde

Academiejaar 2014-2015

Redactie: Filip Verroens

M.m.v. Kristof Baten, Klaas Bentein, Evie Coussé, Claudia Crocco, Lieven Danckaert, Bernard De Clerck, Tine Defour, Annemie Demol, Renata Enghels, Anne-Sophie Ghyselen, Peter Lauwers, Gudrun Rawoens, Miriam Taverniers, Roxane Vandenberghe, Marleen Van Peteghem.

## Inhoud

Woord vooraf.....	4
-------------------	---

Hoofdstuk 1 Heuristiek.....	5
-----------------------------	---

1.1 Publicaties zoeken .....	5
1.1.1 Boeken & tijdschriften.....	5
1.1.2 Bibliografieën .....	6
1.1.3 Andere bronnen .....	7
1.1.4 Zoekacties automatiseren .....	8
1.1.5 Overzicht.....	9
1.2 Referenties beheren.....	9
1.2.1 Beheer via online tools.....	9
1.2.2 Beheer via standalone software.....	10
1.2.3 Het gebruik van Endnote .....	10
1.2.4 Demonstratieoefening Endnote.....	10
1.3 Referenties correct gebruiken .....	15
1.3.1 Soorten publicaties .....	15
1.3.2 Citeren en verwijzen in taalkundige papers .....	16
1.3.3 Plagiaat.....	16

Hoofdstuk 2 Kwantitatieve onderzoeksmethodes .....	20
--	----

2.1 Gegevens verzamelen : Overzicht van de corpora voor de gedoceerde talen binnen de vakgroep Taalkunde.....	20
2.1.1 Duits .....	20
2.1.2 Engels .....	20
2.1.3 Frans .....	24
2.1.4 Italiaans .....	28
2.1.5 Klassieke talen.....	31
2.1.6 Nederlands .....	32
2.1.7 Portugees .....	34
2.1.8 Scandinavistiek .....	35
2.1.9 Spaans .....	38
2.1.10 Het internet als corpus.....	40
2.1.11 Parallele corpora .....	41
2.1.12 Learner corpora .....	42

2.2	Gegevens selecteren: Het gebruik van concordancers .....	42
2.2.1	WordSmith Tools 6.0: WordList functie .....	43
2.2.2	WordSmith Tools 6.0: Concord functie.....	44
2.3	Gegevens beheren: Het gebruik van een database .....	46
2.3.1	Inleiding tot Acces .....	47
2.3.2	Tabel ontwerpen .....	48
2.3.3	Gegevens importeren .....	54
2.3.4	Formulier ontwerpen .....	55
2.3.5	Queries ontwerpen .....	60
2.3.6	Resultaten exporteren.....	75
2.4	Gegevens samenvatten met draaitabellen .....	76
<b>Terminologie .....</b>		<b>82</b>
<b>Bibliografie.....</b>		<b>83</b>
<b>Bijlagen .....</b>		<b>85</b>
3.1	Bijlage A: Criteria voor de Bachelor-/ Masterproef.....	85
3.2	Bijlage B: Grammaticale terminologie .....	86

## Woord vooraf

Dit vademecum richt zich tot de student maar ook tot elke (beginnende) vorser in het domein van de Taalkunde. Het document bevat ongetwijfeld lacunes en grijze zones. Zo is de inventaris van de corpora helemaal niet exhaustief, maar we hopen tenminste de belangrijkste vermeld te hebben. Soms is de info te expliciet, elders is ze het misschien te weinig. Daarom zijn alle opmerkingen die ertoe kunnen bijdragen om deze handleiding te verbeteren zeer welkom! Mail je opmerkingen of suggesties naar [filip.verroens@ugent.be](mailto:filip.verroens@ugent.be).

Editie 1: 2011-2012

Editie 2: 2012-2013      § 2.4 toegevoegd

Editie 3: 2013-2014      § 2.2 herwerkt

Editie 4: 2014-2015      § 1.2.4 toegevoegd

## Hoofdstuk 1 Heuristiek

### 1.1 Publicaties zoeken<sup>1</sup>

In deze sectie gaan we na wat er gepubliceerd is over het onderwerp (bibliografische databanken), waar je de publicaties kan vinden (catalogi) en hoe je ze in je bezit kan krijgen (fulltext online, papier in een UGent bibliotheek, IBL). Bij het opsporen van referenties hanteren we het principe dat we **eerst de meest recente publicaties** over het onderwerp opzoeken alvorens minder recente werken te nemen.

#### 1.1.1 Boeken & tijdschriften<sup>2</sup>

- Voor boeken binnen UGent: <http://lib.ugent.be/> (voor oudere werken eventueel via Collections > Card Catalogue)
- Voor boeken in binnenlandse universiteiten: <http://lib.ugent.be/> > bovenaan tab Databases > Catalogs > **UniCat** : virtual Belgian union catalogue
- Voor tijdschriften in binnenlandse universiteiten: <http://lib.ugent.be/> > bovenaan tab Databases > Catalogs > **Antilope** ([Belgian union catalogue for periodicals.](#))
- Voor (elektronische) tijdschriftartikels: <http://lib.ugent.be/> > naam tijdschrift invullen > links type publicatie selecteren (periodical)

Heel vaak kan je doorklikken naar een elektronische versie van het artikel dat je vervolgens meteen in je Endnote bibliotheek opslaat. Is er enkel een papieren versie beschikbaar, dan

---

<sup>1</sup> We raden aan om eerst de algemene presentatie [introductiesessies](http://lib.ugent.be/demo.shtml) (<http://lib.ugent.be/demo.shtml>) van de UGent te bekijken alsook de site over informatievaardigheden :

<http://www.flw.ugent.be/bibliotheek/diensten/informatievaardigheden>

Tenslotte is de wegwijzer voor bibliotheken & documentatiecentra nu ook online beschikbaar via

<http://search.ugent.be/meercat/x/view/dbs01/002043547>

<sup>2</sup> Merk op dat veel van de hieronder beschreven bronnen slechts gebruikt kunnen worden via het UGent domein. Werk je echter van thuis uit, dan zal eerst een VPN verbinding gemaakt moeten worden naar de UGent. Dit doe je via <http://www.helpdesk.ugent.be/vpn/>.

wordt de precieze UGent bibliotheek vermeld. Mocht het tijdschrift niet beschikbaar zijn aan de UGent, dan worden hier andere binnenlandse instellingen vermeld die het tijdschrift wel bezitten. In dat geval kan je een **IBL** (InterBibliothecair Leenverkeer) aanvraag overwegen (ook buitenland).

## 1.1.2 Bibliografieën

Doelstelling: Referenties van een bepaald onderwerp zoeken of een lijst van publicaties van een bepaalde auteur samenstellen.

### 1.1.2.1 Typologie

- Indeling volgens tijd:
  - Lopende bibliografie: Verschijnt jaarlijks. Geeft de publicaties van één jaar weer.
  - Retrospectieve/ afgesloten bibliografie: Verschijnt één keer. Geeft alle eerdere publicaties weer.
- Indeling volgens domein:
  - Algemene/ universele bibliografie: algemene publicaties uit een bepaalde discipline.
  - Speciale of vakbibliografie: bepaald aspect uit een discipline.
- Indeling volgens inhoud:
  - Signalement bibliografie: geeft slechts het signalement weer.
  - Kritische bibliografie: geeft eveneens een kritisch commentaar van de genoemde publicaties.

### 1.1.2.2 Enkele voorbeelden

- A Bibliography of Tense, Verbal Aspect, Aktionsart and Related Areas (<http://www.utoronto.ca/~binnick/old%20tense/List1.pdf>) [type: retrospectief, speciaal, signalement]
- Bal, W., J. Germain & J.-R. Klein (1991). Bibliographie sélective de linguistique romane et française. Louvain-la-Neuve: Duculot. 268. [type: retrospectief, speciaal, signalement]
- **BALF** (Bulletin analytique de linguistique française)
  - Periode: 1969-
  - Type: lopend, speciaal (Franse taalkunde), kritisch
  - Locatie: L14B Franse Taalkunde
- **BL** (Bibliographie linguistique de l'année ... et complément des années précédentes/ Linguistic bibliography for the year ... and supplement for previous years )
  - Periode: (1939/47) –
  - Type: lopend, algemeen, signalement
  - Locatie<sup>3</sup>: o.a. L14D Romaanse Taalkunde en L18 ATW (tot 1994)

---

<sup>3</sup> De laatste jaargangen kunnen, mits een paswoord, ook online geraadpleegd worden op <http://www.linguisticbibliography.com>.

- **BLL** (Bibliographie Linguistischer Literatur)
  - Periode: 4(1978)-
  - Type: lopend, algemeen, signalement
  - Locatie: centrale bibliotheek - tijdschriftenleeszaal ("BTSL")
- **Francis**
  - Type: lopend, algemeen, signalement. Bevat ruim 1,6 miljoen referenties van artikels, congresverslagen, onderzoeksrapporten en thesissen uit 4.000 tijdschriften over kunstgeschiedenis, kunst en archeologie, filosofie, geschiedenis van literatuurwetenschap, godsdienstwetenschap, pre- en protohistorie, linguïstiek, onderwijs, volkenkunde, enz. met publicaties in het Duits, Engels, Frans, Italiaans en Spaans. Jaarlijks worden ruim 60.000 nieuwe records toegevoegd. Gebaseerd op de *Bulletin Signalétique* van het CNRS.
  - Locatie: <http://lib.ugent.be/> > Databanken > Zoek: Francis
- **LLBA** (*Linguistics and Language Behaviour Abstracts*)
  - Type: algemeen, signalement. Meer dan 361.000 bibliografische records met abstracts uit tijdschriften, (hoofdstukken van) boeken, eindverhandelingen, boekbesprekingen op gebied van linguïstiek en taal (uitspraak, betekenis en woordstructuur, taalvormen, spelling, klankleer, enz.). Ruim 1500 tijdschriften worden geïndexeerd vanaf 1973. Jaarlijkse aangroei met ongeveer 14.000 nieuwe records.
  - Locatie: <http://lib.ugent.be/> > Databanken > Zoek: LLBA<sup>4</sup>
- **MLA** (*Modern Language Association of America*. International bibliography of books and articles on the modern languages and literatures)
  - Type: algemeen, signalement. bevat ruim 2,3 miljoen bibliografische referenties uit ruim 4400 tijdschriften, reeksen en monografieën betreffende linguïstiek, moderne talen, literatuur en folklore. Begint in 1926 tot heden.
  - Locatie: <http://lib.ugent.be/> > Databanken > Zoek: MLA

### 1.1.3 Andere bronnen

- Atypon Link (<http://www.atypon-link.com/>)
- Databanken <http://lib.ugent.be/> > Databanken > Databanken per discipline > L&W bevat m.b.t. taalkunde de volgende links:
  - Academic Search Elite
  - Francis
  - JSTOR (<http://www.jstor.org/>)
  - Linguistics and Language Behavior Abstracts (LLBA)
  - MLA Directory of Periodicals

---

<sup>4</sup> Bij de zoekopdracht in LLBA is er de mogelijkheid om *Peer reviewed* aan te kruisen. **Peer reviewing** wil zeggen dat de publicaties (anoniem) gereviseerd zijn door een leescomité van experts.

- MLA International Bibliography
  - Project Muse (<http://muse.jhu.edu/>)
  - Scirus (<http://www.scirus.com/>)
- 
- E-Books: <http://lib.ugent.be/> > e-boeken via google books of google scholar
  - ERIC (<http://www.eric.ed.gov/>)
  - InformaWorld (<http://www.tandfonline.com/>)
  - <http://www.ingentaconnect.com/>
  - Web of Science (WoS): <http://lib.ugent.be/> > Databases > WoS
  - Elsevier ScienceDirect (<http://www.sciencedirect.com/>)
  - MetaPress (<http://www.metapress.com/home/main.mpx>)

### 1.1.4 Zoekacties automatiseren

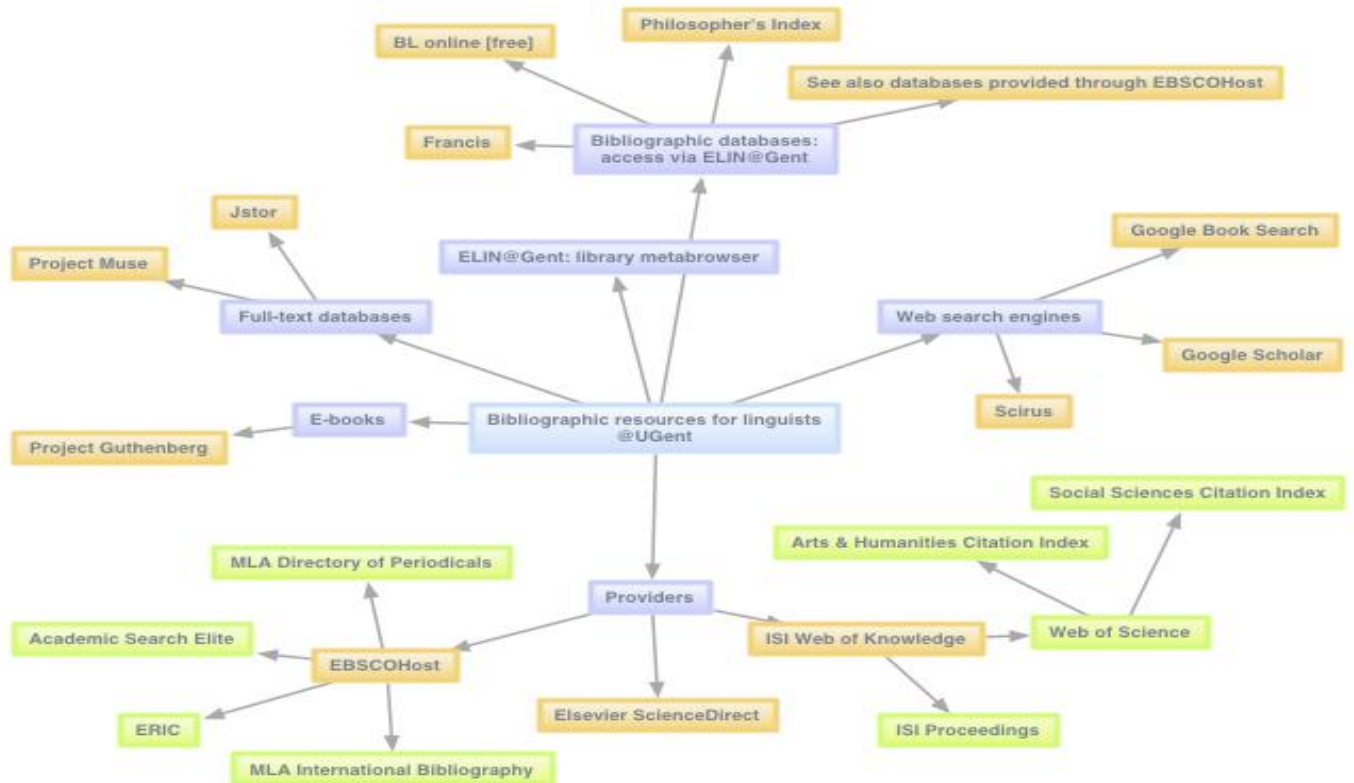
Om informatie te personaliseren en zoekacties te automatiseren kan er gebruik gemaakt worden van attenderingen. Op die manier hoef je niet voortdurend zelf na te gaan of er nieuwe publicaties verschenen zijn over je onderwerp. Naast de **TOC-alert**, attenderingen van nieuwe jaargangen van tijdschriften, is er ook nog de **search alert** en de **citation alert**. De search alert zijn attenderingen op nieuwe records die beantwoorden aan je eerder ingevoerde zoekactie. De **citation alert** zijn attenderingen op nieuwe citaties van publicaties die door jou voorafgaand beschreven zijn. De verschillende attenderingen gebeuren via e-mail of via RSS, m.a.w. je krijgt dan een bericht in je feedreader (Google Reader, Internet Explorer, MS Outlook, etc.). De attenderingen kunnen geactiveerd worden in o.a. [JSTOR](#), [Sage Online Journals](#), [Taylor and Francis Online](#), [Web of Science](#), [Google Scholar](#), [Project Muse](#). Naast deze databanken bestaan er ook TOC-services, d.w.z. gratis portaalsites die de tijdschriften van honderden wetenschappelijke uitgevers verenigen. Hier kan dus een TOC-alert ingesteld worden over de verschillende databanken heen. De belangrijkste voorbeelden zijn [ticTOCs](#) (>14.000 tijdschriften) en [Journal TOCs](#) (>16.000 tijdschriften).



## 1.1.5 Overzicht

### Bibliographic resources for linguists @UGent

M. Taverniers 2008-04-07



Figuur 1: Overzicht bibliografische bronnen

## 1.2 Referenties beheren

### 1.2.1 Beheer via online tools

- Citeulike: [www.citeulike.org](http://www.citeulike.org)
- Connotea: [www.connotea.org](http://www.connotea.org)
- Endnote web: [myendnoteweb.com/](http://myendnoteweb.com/) (vrij voor users van ISI Web of Knowledge)
- RefWorks: [www.refworks.com](http://www.refworks.com)
- Mendeley: [www.mendeley.com/](http://www.mendeley.com/)
- Zotero: <http://www.zotero.org/>

## 1.2.2 Beheer via standalone software

- Bibdesk: [bibdesk.sourceforge.net](http://bibdesk.sourceforge.net)
- Bibliographix: [home.mybibliographix.com](http://home.mybibliographix.com)
- Bookends: By [Sonny Software](http://Sonny Software)
- Endnote: vrij beschikbaar via Athena
- Procite: [www.procite.com](http://www.procite.com)
- Reference Manager: vrij beschikbaar via Athena

## 1.2.3 Het gebruik van Endnote

Endnote (EN) maakt o.a. het volgende mogelijk:

- bijhouden van een database (ook wel bibliotheek genoemd) met referenties
- importeren van referenties afkomstig van andere databases
- referenties uit de database in een tekstdocument integreren
- produceren van bibliografieën in de gewenste stijl

Op <http://www.endnote.com> vind je Engelstalige handleidingen en de allerlaatste connectiefiles, import filters en output stijlen. Nederlandstalige handleiding van EN X7 vind je op <http://www.ugent.be/bw/nl/diensten/bibliotheek/endnote7.pdf>.

## 1.2.4 Demonstratieoefening Endnote<sup>5</sup>

Met behulp van deze oefening leren jullie (1) een eigen referentiebibliotheek op te bouwen, (2) via die bibliotheek referenties in te voegen in Word en (3) die referenties te bewerken en aan te passen naar de gewenste stijl.

### Stap 1. Het programma openen

1. VPN-verbinding maken (niet nodig op Ugent-computer)
2. Start Athena ([athena.ugent.be](http://athena.ugent.be))
3. Open de map 'Office'
4. Selecteer Endnote

### Stap 2. Referentiebibliotheek aanmaken

1. New → Geef **naam** aan nieuwe bibliotheek

---

<sup>5</sup> Red.: Anne-Sophie Ghyselen & Roxane Vandenberghe, februari 2014.

2. Selecteer een locatie waar je je bibliotheek wil opslaan. Als je via Athena werkt, kun je je bibliotheken enkel opslaan op je netwerkschijf (H-schijf)<sup>6</sup> of op de C-schijf van de lokale computer. Wil je je data bewaren op een memorystick, dan dien je die achteraf manueel te kopiëren van de H- of C-schijf.

### **Stap 3. Referenties toevoegen**

In je nieuw aangemaakte bibliotheek moet je nu ook referenties toevoegen. Referenties toevoegen kan op **drie manieren**.

#### 1. Met de hand

**References → new reference → selecteer het juiste “reference type” (bv. tijdschriftartikel, boek) → Vul lege velden in (wees daarin consistent!) → afsluiten + opslaan**

Tips:

- Om een auteur/redacteur in te geven, heb je twee opties: “Achternaam, Voornaam/Voorletter” of “Voornaam/Voorletter achternaam”.
- In je Endnote-bibliotheek zelf hoeft je geen rekening te houden met lay-out of ‘stijl’; namen van auteurs plaats je dus niet in hoofdletters (dus niet: PLEVOETS, maar wel: Plevoets) en titels van boeken niet in cursief.
- Bij meerdere auteurs/redacteurs geef je alle auteurs/redacteurs onder elkaar in. Met andere woorden: geen komma’s, “en”, “&” of “and”!
- Een hoofdstuk/artikel in een boek hoort onder het label “book section”.

**Oefening:** geef volgende referenties met de hand in:

- Boek : ABSILLIS K., J. JASPERS & S. VAN HOOFF (red.). (2012). *De manke usurpator : over verkavelingsvlaams*. Gent : Acedemia Press.
- Artikel uit tijdschrift : DE CALUWE, J. (2009). Tussentaal wordt omgangstaal in Vlaanderen. *Nederlandse Taalkunde* 14-1: 8-25.
- Artikel uit online tijdschrift: PLEVOETS, K. (2009). Verkavelingsvlaams als de voertaal van de verburgerlijking van Vlaanderen. *Studies van de Belgische Kring voor Linguïstiek* 4. <[www.bkl-cbl.be](http://www.bkl-cbl.be)> (02/02/2012).
- Artikel uit boek: RYS, K. & TAELDEMAN, J. (2007). Fonologische ingrediënten van Vlaamse tussentaal. In: D. Sandra, R.Rymenans, P. Cuvelier, & P. Van Petegem (red.).

---

<sup>6</sup> Aan die schijf kun je thuis ook via je VPN-verbinding, maar dan moet je de schijf wel ‘mounten’ ([http://helpdesk.ugent.be/netdisk/bestand\\_mount.php](http://helpdesk.ugent.be/netdisk/bestand_mount.php)).

*Tussen taal, spelling en onderwijs. Essays bij het emeritaat van Frans Daems.* Gent: Academia Press: 23-34.

Zoek het artikel van Plevoets ook online op, sla het op op je computer in pdf en voeg het toe aan de referentie: **open de referentie in endnote → klik op de rechtermuistoets of klik icoontje (paperclip) rechts boven aan → File Attachment → Attach File → zoek op je PC het gewenste (pdf-)bestand en dubbelklik erop.**

## 2. Vanuit online database

Je kunt ook een referentie toevoegen aan je Endnotebibliotheek vanuit een **bibliografische databank**, zoals de MLA en WoS. Voordelen: (1) Heel gedetailleerde referenties, (2) heel snel, (3) geen typfouten.

**Oefening:** zoek in de MLA een artikel van Johan Taeldeman over “grammaticale kenmerken van tussentaal” en voer de referentie in in Endnote.

- Open **via Athena (belangrijk!)** een **internetbrowser** (Internet Explorer)
- Open via die browser de MLA Databank: **lib.ugent.be → Databanken → Typ “MLA” → “Start Database”**
- Artikel zoeken: bvb. trefwoorden “tussentaal” en “Taeldeman”
- Referentie exporteren: **Selecteer artikel (vink het aan!) → “Export/Save” (rechts) → selecteer “Endnote/Referencemanager” → continue → Selecteer onderaan “open” → selecteer “Endnote” → OK**
- Referentie verschijnt automatisch in je Endnote library

## 3. Vanuit iemand anders’ bibliotheek

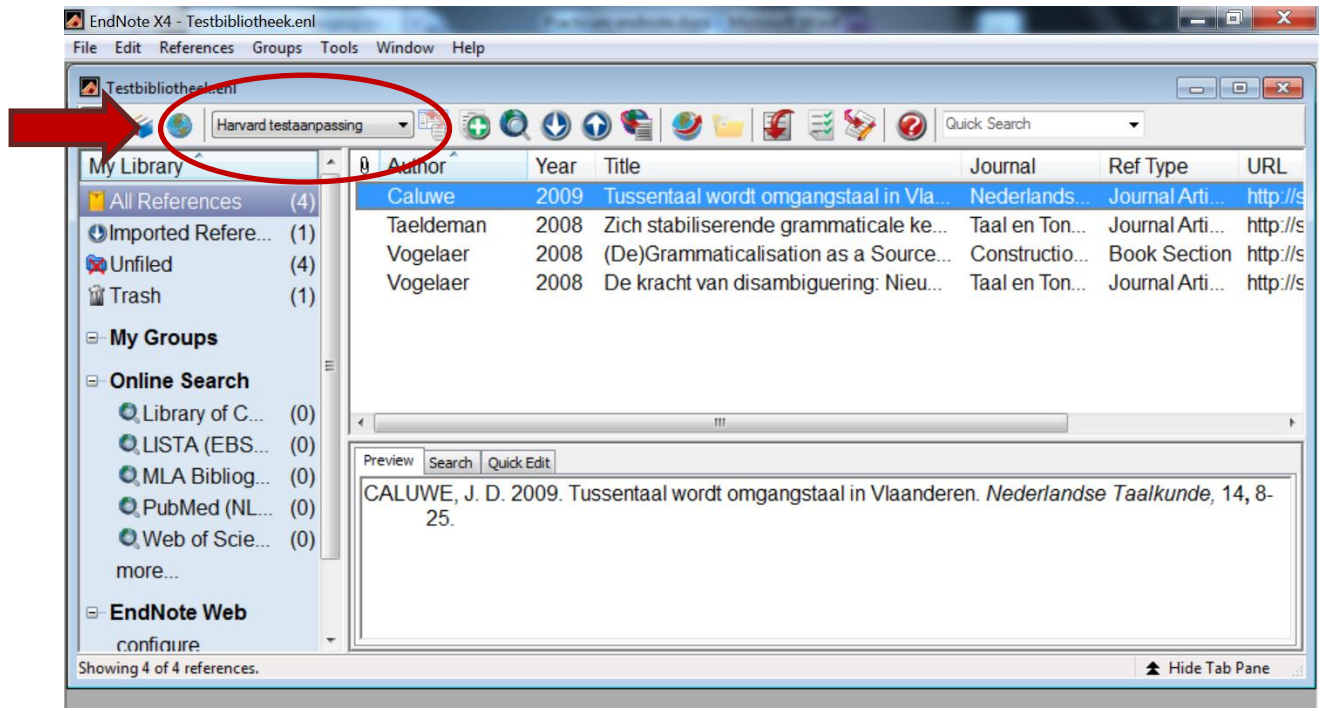
Handig aan Endnote is dat je andermans bibliotheek (bvb. die van een medestudent die rond hetzelfde thema werkt) perfect aan de jouwe kunt toevoegen. Dat doe je als volgt:

- Sla de bibliotheek (met bijhorende datamap) die je aan je eigen bib wil toevoegen op op je computer of netwerkschijf
- Open de Endnote library waar je al je referenties in wil stoppen → **File → Import → “Choose File” (zoek bibliotheek die je wilt toevoegen) → “Import Options”: “Endnote Library” → “Duplicates”: “discard any duplicates” (om te vermijden dat je verschillende referenties dubbel hebt) → Import**

## Stap 4. Stijl selecteren en aanpassen

Voor je je tekst begint te schrijven, selecteer je het best de **stijl** waarin je wenst te verwijzen en pas je die aan naar je eigen wensen (of die van de lesgever of het tijdschrift waarvoor je een paper schrijft).

## 1. Stijl selecteren



Endnote biedt een honderden stijlen aan waaruit je kunt kiezen, zoals de Harvardstijl, de MLA-stijl en de stijl van Journal of Linguistics.

**Select another style > Search of scrollen (bv. *J Linguistics*) > Choose.**

**Tip** : als je op een referentie gaat staan en je klikt rechts op **preview**, dan zie je meteen hoe de referentie in de bibliografie weergegeven zal worden.

## 2. Stijl aanpassen

Ondanks het enorme gamma aan stijlen is het toch vaak moeilijk om de stijl te vinden die perfect aan jouw behoeften beantwoordt. Gelukkig kunnen de bestaande stijlen aangepast worden en opgeslagen als een nieuwe, ‘eigen’ stijl. Dat doe je door eerst **de stijl te selecteren die je wil aanpassen, bv. *J Linguistics* → “Edit” → “Output style” → Edit *J Linguistics*.**

Zowel voor de referenties in de tekst (“citations”), als voor de bibliografie (“bibliography”) vind je onder **Templates** alle kenmerken van die bepaalde stijl op een rijtje, en die kun je naar believen aanpassen. Wanneer je vervolgens op afsluiten klikt, dan kun je de veranderingen in de stijl opslaan onder een nieuwe stijlnaam. Die nieuwe stijl kun je achteraf nog aanpassen (**Stijl selecteren → “Edit” → “Output style” → “Edit ‘aangepaste stijl’**”).

**Oefening**: in de *J Linguistics*-stijl staan de paginanummers van een artikel uit een boek (*book section*) vòòr de plaats en de uitgeverij in de bibliografie. Verander de stijl zò dat de

paginanummers pas na de uitgeverij worden geplaatst, gescheiden door een dubbele punt. Ga daarna terug naar je hoofdtabblad en selecteer nu de aangepaste J Linguistics-stijl als defaultstijl voor al je referenties.

### **Stap 5. Citeren in tekst** (“cite while you write”)

Nu je een bibliotheek aangelegd hebt en je stijl hebt gedefinieerd, kun je aan de slag in Word (**openen via Athena !! → Office**). Daar vind je een tabblad “**EndNote**”, waarmee je verwijzingen kunt invoegen (**insert citation**). Wanneer je een citaat invoegt, krijg je standaard de weergave “(AUTEUR JAAR)” en op het einde van je document de bibliografische referentie. Vaak hebben we in onze tekst echter ook paginaverwijzingen nodig in de referentie of willen we de auteur of het jaartal niet binnen de haakjes te vermelden omdat die al in de zin vermeld staat. Dat kun je aanpassen door de referentie te selecteren en aanpassingen aan te geven → **rechtermuisknop** → “**Edit citation**” → **verschillende opties** : **jaartal/auteur weglaten of “More”** : **paginanummers toevoegen**.

The screenshot shows the 'EndNote X4 Edit & Manage Citations' dialog box. It has two tabs: 'Edit Citation' and 'Reference'. The 'Edit Citation' tab is active, showing fields for 'Prefix', 'Suffix', and 'Pages'. There are checkboxes for 'Exclude author' and 'Exclude year'. Red arrows point from text boxes to these options:

- Arrow to 'Exclude author': **Hiermee verwijder je de auteur uit de referentie**
- Arrow to 'Exclude year': **Hiermee verwijder je het jaartal uit de referentie**
- Arrow to 'Prefix': **Hiermee kun je prefixen toevoegen aan je referentie, zoals in de referentie (zie bijvoorbeeld Taeldeman 2011)**
- Arrow to 'Suffix': **Hiermee kun je suffixen toevoegen aan je referentie, zoals in de referentie (Plevoets 2011 noemt dit fenomeen)**
- Arrow to 'Pages': **Hiermee voeg je pagina's toe aan je referentie**

**Oefening:** Typ de volgende tekst in in Word en voeg daarbij telkens de juiste referentie in.

*Wie er de literatuur op naslaat, zal gauw merken dat er in de literatuur over tussentaal in de voorbije decennia een aanzienlijke klemtoonverandering heeft plaatsgevonden. Terwijl*

vroeger vooral hevig oplopende opiniestukken over de aanvaardbaarheid van tussentaal werden gepubliceerd, verschijnen de laatste jaren veel vaker empirische studies van tussentaal. Rys & Taeldeman publiceerden in 2007 bijvoorbeeld een empirische studie over de fonologische ingrediënten van tussentaal, waarin ze bepleiten dat hoewel momenteel vooral variatie troef is, er zich toch een stabiliserende tussentaal aan het vormen is (cfr. Taeldeman 2008). Die these wordt echter tegengesproken door de Plevoets (2008:175), die bij zijn corpusonderzoek naar morfologische variatie in Vlaanderen vaststelde dat standaardtaal en tussentaal “een naadloze overgang [vertonen] van en naar elkaar”.

## 1.3 Referenties correct gebruiken

### 1.3.1 Soorten publicaties

We kunnen verschillende soorten publicaties onderscheiden (monografie/ boek, tijdschriftartikels, hoofdstukken in boeken, (thematische) volumes, congresverslagen (*conference proceedings*), encyclopedieën, woordenboeken, dissertaties, reviews, abstracts, squibs<sup>7</sup>, etc.). Elk type publicatie stemt overeen met een bepaald soort bibliografische notatie (cf. 1.3.2). Interessant om weten is ook dat de verschillende soorten publicaties ingedeeld worden in categorieën (zie codes onder). De code zegt dus iets meer over het type publicatie. Je vindt ze terug in de academische bibliografieën van de UGent onderzoekers.

#### Artikels

- a1 artikels in tijdschriften, opgenomen in ISI Web of Science<sup>8</sup> – beperkt tot artikels van het type: article, review, letter, note
- a2 artikels in ruim verspreide wetenschappelijke tijdschriften
- a3 artikels in nationale tijdschriften met leescomité niet begrepen in (a1) of (a2),
- a4 artikels in tijdschriften niet begrepen in (a1), (a2) of (a3).

#### Boeken

- b1 auteur of co-auteur van boeken (beperkt tot boeken uitgegeven bij een wetenschappelijke uitgeverij)
- b2 hoofdstukken in boeken

---

<sup>7</sup>Squibs zijn doorgaans erg korte stukjes die bedoeld zijn om na te denken en discussie uit te lokken over onderwerpen van (theoretisch) belang. Maar ze kunnen ook de bedoeling hebben om afwijkende gegevens naar voren te schuiven zonder daarbij een oplossing te bieden.

<sup>8</sup>Binnen de A1 tijdschriften kan er nog een onderscheid gemaakt worden volgens de **Impact Factor**, een factor die bepaald wordt op basis van de bibliometrische gegevens van Web of Science. Een Impact factor is het gemiddelde aantal citaties naar artikelen van een tijdschrift gedurende de afgelopen twee jaar. Impact factoren zijn sterk afhankelijk van de wetenschappelijke discipline. Organisaties als [CWTS](#) proberen door normalisatie van de gegevens toch tot vergelijkingen te komen. Een overzicht van Impact factoren is te vinden in de [Journal Citation Reports](#).



b3 boeken als editor

#### Andere publicaties

- c1 artikels in proceedings van wetenschappelijke congressen, niet inbegrepen in vorige rubrieken (volledige artikels met uitsluiting van abstracts)
- c2 proefschriften, verhandelingen, interne verslagen, abstracts van mededelingen op congressen.
- c3 octrooien

#### Dissertaties

- d1 Phd proefschrift
- d2 Master proefschrift

### 1.3.2 Citeren en verwijzen in taalkundige papers

Er kan op verschillende manieren geciteerd en verwezen worden. In het domein van de taal- en letterkunde zijn de meest courante stijlen de APA (zie [APAStyle.org](http://APAStyle.org)) en de MLA (zie [summary](#) of the MLA citation style). Voor de taalkunde verschilt de stijlopmaak vaak per tijdschrift bv. bij *Journal of Linguistics* (Cambridge), *Language* (LSA), *Linguistic Inquiry* (MIT Press) en *Stylesheet for De Gruyter Mouton journals*. In de context van de onderzoekstaak verwijzen we hier naar de overzichtelijke handleiding van Muriel Norde (2007) en vooral *The Generic Style Rules for Linguistics* (Haspelmath 2014) die meer en meer internationaal toegepast worden. Deze laatste liggen in de lijn van de Unified Style Sheet (°2007) en integreren *The Leipzig Glossing Rules for interlinear morpheme-by-morpheme glosses*, interessant als men voorbeelden moet glosseren. **Het is van het hoogste belang dat het citeren en verwijzen op een consistente manier gebeurt** voor artikels, boeken, hoofdstukken in boeken, etc. Het gebruik van software (bv. Endnote) vermindert het aantal fouten aanzienlijk. Sommige stijlen gebruiken meer gegevens dan andere (bv. vermelding van de volledige voornaam i.p.v. enkel de initialen). Het is dus belangrijk dat je zoveel mogelijk gegevens noteert om te vermijden dat je nadien nog extra gegevens van referenties moet opzoeken (bv. uitgever, pagina's, etc.).

### 1.3.3 Plagiaat<sup>9</sup>

#### 1.3.3.1 Definitie UGent

Besluit betreffende de definitie van plagiaat in uitvoering van artikel 79, § 2 van het Onderwijs- en Examenreglement 2009-2010

---

<sup>9</sup> Bron: <http://www.flwi.ugent.be/plagiaat>



§2. Het plegen van plagiaat is een vorm van fraude. De examencommissie per opleiding legt in een sluitende definitie het begrip "plagiaat" voor die opleiding vast. Deze informatie wordt aan de studenten van de opleiding gecommuniceerd. De verantwoordelijke evaluator kan de originaliteit van ingediende werkstukken, die bovendien steeds in elektronische vorm moeten worden ingediend, middels anti-plagiaatsoftware toetsen.

Indien een verantwoordelijk lesgever vermoedt dat een student plagiaat heeft gepleegd dat de evaluatie van het betreffende werkstuk kan beïnvloeden, wordt de voorzitter van de examencommissie per opleiding hiervan onverwijld op de hoogte gebracht.

De examencommissie van elke opleiding van de faculteit Letteren en Wijsbegeerte is overeengekomen om een voor de hele faculteit gemeenschappelijke definitie van plagiaat vast te leggen. De definitie van plagiaat wordt als volgt vastgelegd:

Het overnemen, al dan niet in vertaling, van andermans woorden of het parafraseren van iemands gedachtegang zonder de vindplaats te vermelden of zonder aan te geven wiens ideeën worden overgenomen, wordt gezien als plagiaat. Alle stukken tekst (zinnen of zinsneden) die men overneemt uit om het even welke andere publicatie en om het even welke taal, zelfs uit webpagina's op het internet of uit een manuscript, moeten dus een bronvermelding krijgen en – wanneer letterlijk overgenomen – als citaat gekenmerkt zijn. Een globale bronvermelding, bijv. bij het begin van een hoofdstuk, verschoont het plagiaat t.o.v. afzonderlijke passages niet. Het vermelden van een bron in de bibliografie of bronnenoverzicht verschoont het plagiaat niet.

Enkele voorbeelden

**Voorbeeld 1:**

Oorspronkelijke tekst: "Wanneer politieke partijen collectieve belangen verdedigen, dan mag men er zeker van zijn dat de fundamentele maatschappelijke problemen op de politieke agenda worden geplaatst. Door collectieve belangen te verdedigen, beschikken zij over een natuurlijk electoraat (de dragers van die belangen), aangevuld met kiezers die zich hebben laten overtuigen."

Deze tekst heeft als bron: P. Saey, "De Opkomst van Extreem-Rechts als Geografisch probleem", in D. Praet (red), *"Us and Them". Essays over filosofie, politiek, religie en cultuur van de Klassieke Oudheid tot Islam in Europa. Ter ere van Herman De Ley*, Gent, Academia Press, 2008, p. 359-360.

→ Als je het idee dat hier wordt verwoord, wenst te gebruiken in je eigen paper, kan je van oordeel zijn dat je de tekst best volledig overneemt. Dan moet je aangeven dat het om een citaat gaat: je moet de tekst tussen aanhalingstekens plaatsen en de bron vermelden. Doe je dit niet, dan is er sprake van plagiaat.

**Geen plagiaat:**

"Wanneer politieke partijen collectieve belangen verdedigen, dan mag men er zeker van zijn dat de fundamentele maatschappelijke problemen op de politieke agenda worden geplaatst. Door collectieve belangen te verdedigen, beschikken zij over een natuurlijk electoraat (de dragers van die belangen), aangevuld met kiezers die zich hebben laten overtuigen." (P. Saey, 2008, p. 359-360)

**Wel plagiaat:**

Wanneer politieke partijen collectieve belangen verdedigen, dan mag men er zeker van zijn dat de fundamentele maatschappelijke problemen op de politieke agenda worden geplaatst. Door collectieve belangen te verdedigen, beschikken zij over een natuurlijk electoraat (de dragers van die belangen), aangevuld met kiezers die zich hebben laten overtuigen. (P. Saey, 2008, p. 359-360)

**Wel plagiaat:**

Wanneer politieke partijen collectieve belangen verdedigen, dan mag men er zeker van zijn dat de fundamentele maatschappelijke problemen op de politieke agenda worden geplaatst. Door collectieve belangen te verdedigen, beschikken zij over een natuurlijk electoraat (de dragers van die belangen), aangevuld met kiezers die zich hebben laten overtuigen.

→ Als je het idee dat hier wordt verwoord, in eigen bewoordingen wenst te gebruiken in je eigen paper, dan moet je ook aangeven waar je het idee vandaan hebt. Doe je dit niet, dan is er sprake van plagiaat.

**Geen plagiaat:**

Fundamentele maatschappelijke problemen zullen maar op de politieke agenda worden geplaatst, wanneer politieke partijen collectieve belangen verdedigen, want dan beschikken ze over een natuurlijke groep van kiezers, namelijk zij die voor deze belangen staan. (P. Saey, 2008, p. 359-360)

**Wel plagiaat:**

Fundamentele maatschappelijke problemen zullen maar op de politieke agenda worden geplaatst, wanneer politieke partijen collectieve belangen verdedigen, want dan beschikken ze over een natuurlijke groep van kiezers, namelijk zij die voor deze belangen staan.

**Voorbeeld 2:**

Oorspronkelijke tekst: “Dutch colonial rule proclaimed that different groups required different legislation; this resulted in the legal invention of second class citizenship. When at the turn of the century, Dutch women advocated citizenship for women, they could identify partially with the indigenous population of the colonies. This however, did not result in real solidarity between Dutch and Indonesian feminists.”

Deze tekst heeft als bron: B. Waaldijk, “Subjects and Citizens: Gender and Racial Discrimination in Dutch Colonialism at the End of the 19th Century”, in G. Háfðanarson (ed.), *Racial Discrimination and Ethnicity in European History*, Pisa, Edizioni Plus – Università di Pisa, 2003, p.111.

→ Een letterlijke vertaling in het Nederlands van een anderstalige tekst, is ook plagiaat. Je moet bij vertalingen op een behoorlijke wijze aan de oorspronkelijke tekst refereren.

**Geen plagiaat:**

Volgens Waaldijk vond het Nederlandse koloniale regime dat een verschillende wetgeving vereist is voor verschillende groepen, waardoor een zogenaamd tweede klasse burgerschap ontstond. Toen, bij de eeuwwisseling, de Nederlandse vrouwen pleitten voor burgerschap voor vrouwen, konden ze zich gedeeltelijk met de inheemse

bevolking identificeren. Dit resulteerde nochtans niet in echte solidariteit tussen Nederlandse en Indonesische feministes. (B. Waaldijk, 2003, p.111)

**Wel plagiaat:**

Het Nederlandse koloniale regime bepaalde dat verschillende wetgeving vereist is voor verschillende groepen; dit resulteerde in de wettelijke uitvinding van tweede klassen burgerschap. Toen, bij de eeuwwisseling, de Nederlandse vrouwen pleitten voor burgerschap voor vrouwen, konden zij zich gedeeltelijk met de inheemse bevolking van de kolonies identificeren. Dit resulteerde nochtans niet in echte solidariteit tussen Nederlandse en Indonesische feministes.

Als je je kennis van plagiaat wil testen, kan je terecht op de website van Indiana University Bloomington, zie: <http://www.indiana.edu/~tedfrick/plagiarism/>

## Hoofdstuk 2 Kwantitatieve onderzoeksmethodes

### 2.1 Gegevens verzamelen : Overzicht van de corpora voor de gedoceede talen binnen de vakgroep Taalkunde

#### 2.1.1 Duits

##### 2.1.1.1 Diachrone corpora

##### 2.1.1.2 Synchrone corpora

##### 2.1.1.2.1 Gesproken corpora

##### 2.1.1.2.2 Geschreven corpora

#### ➤ COSMAS II (Corpus Search, Management and Analysis System)

- Periode: 1772-heden
- Grootte: meer als 3,9 Mrd. woorden
- Teksttypes: krantentaal, vaktaal en fictie uit Duitsland, Oostenrijk en Zwitserland
- Nuttig: hét referentiecorpus voor Duits
- Door: Institut für Deutsche Sprache (IDS)
- Informatielink: <http://www.ids-mannheim.de/cosmas2/>
- Toegang: mogelijk na registratie

#### ➤ German Corpus Page (<http://corp.hum.sdu.dk/corpuseye.de.html>)

#### 2.1.2 Engels

##### 2.1.2.1 Diachrone corpora<sup>10</sup>

#### ➤ A Corpus of English Dialogues 1560-1760 (CED)

- Periode: 1560-1760
- Grootte: 1.2 miljoen woorden (1,183,690)
- Genre: Diachroon (vroegmodern Engels)
- Teksttypes: Gesproken interactie: (historische) dialogen. Authentieke dialogen (i.e. ‘witness depositions & trial proceedings’) maar ook gereconstrueerde dialogen in de vorm van fictieteksten of toneelstukken.
- Nuttig: Heel nuttig bij onderzoek naar historische pragmatiek. Het corpus bevat enkel dialogen (i.e. o.a. ‘authentieke’ (gereconstrueerde) dialogen zoals rechtsgetuigenissen en ‘trial proceedings’ – en ‘geconstrueerde’ dialogen

---

<sup>10</sup> Geüpdatete en correcte informatie i.v.m. corpora (o.a. samengesteld door Varieng, Helsinki), is te vinden op: <http://www.helsinki.fi/varieng/CoRD/corpora/>

zoals toneelstukken, fictie) en geeft daarbij een andere kijk op historische teksten. De ‘face-to-face’ interactie die zo onderzocht kan worden (weliswaar via ‘geschreven gesproken teksten’) is erg nuttig voor onderzoek naar politeness, pragmatiek, subjectiviteit en conversatiestructuren.

- Door: Merja Kytö & Jonathan Culpeper; o.a. Uppsala University
- Informatielink: <http://www.helsinki.fi/varieng/CoRD/corpora/CED/index.html>  
[http://www.engelska.uu.se/Forskning/sprakvetenskap/A\\_Corpus\\_of\\_English\\_Dialogues/](http://www.engelska.uu.se/Forskning/sprakvetenskap/A_Corpus_of_English_Dialogues/)
- Beschikbaarheid: Beschikbaar op CD-Rom voor niet-commercieel gebruik via  
Oxford Text Archive:  
<http://ota.ahds.ac.uk/scripts/download.php?otaid=2507>  
De cd-rom zou gratis zijn indien voor niet-commercieel gebruik.  
Kan ook gelezen worden met o.a. Wordsmith

➤ **Corpus of Early English Correspondence (Sampler) (CEECS)**

- Periode: 1418-1680
- Grootte: 23 brievencollecties: 450,000 woorden
- Genre: Diachroon (Middelengels, Vroegmodern Engels)
- Teksttypes: Brieven
- Nuttig: Nuttig voor sociolinguistisch of – sowieso – historisch onderzoek
- Door: Universiteit van Helsinki; oa Minna Nevala & Terttu Nevalainen
- Informatielink: <http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/index.html>  
<http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/ceecs.html>
- Beschikbaarheid: ICAME Corpus collection CD-Rom  
(Zie beschrijving Helsinki Corpus)

➤ **Corpus of Late Modern English Texts (Extended Version) (CLMETEV)**

- Periode: 1710-1920
- Grootte: 15 miljoen woorden
- Genre: Formeel, geschreven, Brits Engels (Laatmodern Engels)
- Teksttypes: Verschillende genres: o.a. brieven, fictie, wetenschappelijke teksten
- Nuttig: voor diachrone studie
- Nadeel: Er zijn wel korte beschrijvingen beschikbaar (o.a. een overzicht bij het corpus zelf), maar geen ‘manual’. Het corpus en meer bepaald de verschillende teksten die gebruikt worden zijn daarom niet zo overzichtelijk. Info vinden over periodes, auteur, genre etc. van individuele teksten houdt soms wat zoekwerk in...
- Door: Universiteit van Leuven: Hendrik Desmet
- Informatielink: <http://www.helsinki.fi/varieng/CoRD/corpora/CLMETEV/index.html>  
<http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/index.html>

- Beschikbaarheid: Vrij beschikbaar

Het corpus is samengesteld uit publieke teksten die afzonderlijk allemaal beschikbaar zijn via verschillende online projecten (vb. Project Gutenberg, Oxford Text Archive, Victorian Women Writers Project). Het corpus kan gedownload worden via:

<http://perswww.kuleuven.be/~u0044428/> (user-id en paswoord beschikbaar via contact met Hendrik Desmet)

## 2.1.2.2 Synchrone corpora

### 2.1.2.2.1 Gesproken corpora

#### ➤ CPSAE - Corpus of Spoken Professional American English

- Periode: 1994-1998
- Grootte: 2 million words (2 sub-corpora, approx. 1 million words each)
- Teksttypes: academic discourse and White House briefings; short interchanges by 400 speakers
- Nuttig: untagged and tagged version
- POS tagging on the basis of the CLAWS tagset developed at Lancaster University
- Door: Michael Barlow at Athelstan and Rice University, Houston/TX
- Informatielink:
- Toegang: vanuit Athelstan

#### ➤ COLT - Bergen Corpus of London Teenage Language

- Periode: 1993
- Grootte: 500,000 words
- Teksttypes: transcripts of spoken language of London teenagers (COLT is part of the BNC)
- Nuttig: POS tagging
- Door: University of Bergen, Norway
- Informatielink:
- Toegang: ICAME CD

#### ➤ LLC London-Lund Corpus of Spoken English

- Periode: 1960s, 1975-81, 1985-88
- Grootte: 500,000 words
- Teksttypes: spoken language (more info)  
based on the Survey of English Usage (SEU, 1959, University College London) and on the Survey of Spoken English (SSE, 1975, Lund University)
- Nuttig: prosodic and discourse annotation
- Door: Randolph Quirk and Sidney Greenbaum at University College London  
Jan Svartvik at Lund University

- Informatielink:
- Toegang: ICAME CD

➤ **MICASE - Michigan Corpus of Academic Spoken English**

- Periode: 1997-heden
- Grootte: 1.7 million words
- Teksttypes: transcripts and audio files of academic speech
- Nuttig: discourse annotation
- Door: R. C. Simpson, S. L. Briggs, J. Ovens, and J. M. Swales at the English Language Institute, University of Michigan
- Informatielink:
- Toegang: vrij

➤ **Gesproken deel van BNC (10m): cf. infra**

➤ **CHILDES (Child Language Data Exchange System)**

- Periode:
- Grootte:
- Teksttypes: dé referentie voor de studie van kindertaal. Maakt deel uit van TalkBank.
- Nuttig: vereist CLAN (en TalkBankViewer) om data te bekijken en te analyseren.
- Door:
- Informatielink: <http://childes.psy.cmu.edu/>
- Toegang: downloadable na aanvraag

#### **2.1.2.2.2 Geschreven corpora**

➤ **British National Corpus (BNC)**

- Periode: late 20<sup>ste</sup> eeuw (i.e. 1980s-1993)
- Grootte: 100 miljoen woorden
- Genre: Hedendaags Engels – Geschreven (90 %) en gesproken (10 %)
- Teksttypes: Geschreven teksten: o.a. geschreven pers (kranten), ‘periodicals’, academische boeken en populaire fictie, brieven, (school) essays. Gesproken teksten: o.a. orthografische transcripties van informele conversaties; conversaties uit verschillende contexten, van formele zakelijke gesprekken tot radioshow's.
- Nuttig: Je kan via de tweede link hieronder (of op het corpus met licentie zelf) ook opzoeken via ‘samples’ doen: vb. 50, 100 of meer samples, selecteerbaar op gesproken of geschreven taal, en subtypes.
- Door: o.a. Mark Davies

- Informatielink: <http://corpus.byu.edu/bnc/> en <http://www.natcorp.ox.ac.uk/>
- Beschikbaarheid: Er zijn verschillende versies van het BNC. Die zijn te bestellen via de hoofdsite (zie informatielink). Maar, je kan ook gratis online opzoeken doen (na eerst te registreren).

#### ➤ **Brown corpus**

- Periode: 1961
- Grootte: 1 miljoen
- Teksttypes: Part of the ICAME corpus collection. Written American English, divided into 2,000-word samples from various genres Deel van de ICAME corpuscollectie. Geschreven Amerikaans-Engels verdeeld in 2.000-woord monsters uit verschillende genres.
- Nuttig: Als pionier belangrijk voor de corpuslinguïstiek. Minutieus samengesteld bestaand uit 15 tekstgenres. Corpus POS tagging. Grammatical parsing. Brown Corpus tagset: <http://www.scs.leeds.ac.uk/amalgam/tagsets/brown.html>  
Getagged by the Penn Treebank project (POS tagging; grammatical parsing). Zie ook de Engelse tegenhanger LOB (Lancaster-Oslo-Bergen Corpus) samengesteld volgens dezelfde principes.
- Door: Brown University
- Informatielink:
- Toegang: CD-rom Engelse Taalkunde

- ICE corpus (<http://ice-corpora.net/ice/index.htm>) [Cd-rom Engelse Taalkunde]
- COCA - Corpus of Contemporary American English (free access)
- Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME)
- Penn-Helsinki Parsed Corpus of Middle English (PPCME2)
- Grammar of Spoken Singapore English Corpus  
(info: [http://www0.hku.hk/english/staff/lisa\\_lim/GSSEC-GroundRules-2009.doc](http://www0.hku.hk/english/staff/lisa_lim/GSSEC-GroundRules-2009.doc))

## 2.1.3 Frans

### 2.1.3.1 Diachrone corpora

- Principaux corpus médiévaux <http://ccfm.ens-lsh.fr/spip.php?rubrique13>
- Bibliotheca augustana (losse teksten) [http://www.hs-augsburg.de/~harsch/gallica/Chronologie/f\\_chrono.html](http://www.hs-augsburg.de/~harsch/gallica/Chronologie/f_chrono.html)
- Bibliothèque Electronique de Lisieux <http://www.bmlisieux.com/> (HTML versies van teksten van eerder secundaire auteurs van eind 19<sup>e</sup> begin 20<sup>e</sup> eeuw)



- Projet Charrette <http://www.princeton.edu/~lancelot/>
- La base du français médiéval (BMF) <http://bfm.ens-lyon.fr/>
- Textes de moyen français à l'ATILF <http://atilf.atilf.fr/dmf.htm>
- Laboratoire de français ancien (LFA)  
<http://www.lib.uchicago.edu/efts/ARTFL/projects/TLA/>  
*De twee laatste zijn geïntegreerd in FRANTEXT (cf. infra)*
- Nouveau Corpus d'Amsterdam (NCA) <http://www.uni-stuttgart.de/lingrom/stein/corpus/#nca>
- The Anglo-Norman On-line Hub (<http://www.anglo-norman.net/>)

### 2.1.3.2 Synchrone corpora

#### 2.1.3.2.1 Gesproken corpora

##### ➤ CORPAIX

- Periode:
- Grootte: Nbre total de token..... 1207587
  - Nbre de codes de transcription.. 74878
  - Nbre de formes lexicales..... 1132709
  - Nbre de formes différentes..... 30770
  - Nbre de fichiers\_corpus ..... 271
  - Taux d'indexation mots/sec..... 1974
- Teksttypes: getranscribeerde dialogen
- Nuttig:
- Door: Université d'Aix-en-Provence
- Informatielink:
- Toegang: CD-rom beschikbaar op secretariaat Frans

##### ➤ ELICOP

- Periode: 1968-1976 (Lancom)
- Grootte: Nom du Corpus.....**elilap** (Etude Linguistique de la Langue Parlée)
  - Nbre total de mots dans corpus d'Orléans..... 4.500.000
  - Nbre total de mots dans corpus de Tours..... 36.000
  - Nbre total de mots dans corpus d'Auvergne..... 177.000
  - Nom du Corpus..... **lancom** (Langue et COMMunication)
  - Nbre total de mots.....160.000

Elilap + lancom = **elicop** (4.873.000)

- Teksttypes: getranscribeerde dialogen
- Nuttig: lemmatisering, woordsoortinformatie (POS tagging), prosodische annotatie
- Door: Katholieke Universiteit Leuven (i.s.m. andere partners)
- Informatielink: <http://bach.arts.kuleuven.be/elicop/>
- Toegang: vrij van licentie

- Corpus de Français Parlé Parisien des années 2000 (CFPP2000) : <http://cfpp2000.univ-paris3.fr/>
- Corpus de référence du français parlé <http://www.up.univ-mrs.fr/veronis/pdf/2004-presentation-crfp.pdf>

#### ➤ **Corpus de la SNCF**

- Periode:
  - Grootte: 84.494 woorden
  - Teksttypes: getranscribeerde dialogen van telefoongesprekken uit het informatiecentrum van het treinstation St-Lazare in Parijs.
  - Nuttig: de doelstelling is de analyse van taalkundig gedrag bij gebruikers van het informatiecentrum
  - Door: project GRECO (Grenoble Campus Ouvert)
  - Informatielink:
  - Toegang: CD-rom beschikbaar op secretariaat Frans
- Inventaris door Paul Cappeau van gesproken corpora: [http://www.culture.gouv.fr/culture/dglf/recherche/corpus\\_parole/Inventaire.pdf](http://www.culture.gouv.fr/culture/dglf/recherche/corpus_parole/Inventaire.pdf)
  - Phonologie du français contemporain <http://www.projet-pfc.net/>
  - Valibel (VARIétés LInguistiques du français en BELgique) <http://valibel.fltr.ucl.ac.be/>

#### **2.1.3.2.2 Geschreven corpora**

- ABU (Association des Bibliophiles Universels) <http://abu.cnam.fr/>
- ARTFL American and French Research on the Treasury of the French Language <http://humanities.uchicago.edu/orgs/ARTFL/>
- Athena Textes français <http://un2sg4.unige.ch/athena/html/francaut.html>

- Corpatext <http://www.lexique.org/public/lisezmoi.corpatext.htm>
- Corpus de Leipzig [http://wortschatz.uni-leipzig.de/ws\\_fra/](http://wortschatz.uni-leipzig.de/ws_fra/)
- Corpuseye (<http://corp.hum.sdu.dk/cqp.fr.html>)
- Electronic Text Collections in Western European Literature  
<http://www.lib.virginia.edu/wess/etexts.html>  
i.h.b.: *Classici Stranieri* online Franse werken.
- Base Textuelle **FRANTEXT**
  - Periode: 1330-2009
  - Grootte: ca. 240 miljoen woorden
  - Teksttypes: literair proza, theaterteksten, gedichten, brieven
  - Nuttig: lemmatisering, woordsoortinformatie (POS tagging)
  - Door: CNRS, ATILF, Université de Nancy
  - Informatielink: <http://www.frantext.fr/> Een handleiding is beschikbaar bij Prof. Peter Lauwers
  - Toegang: Betalende licentie per geregistreerd IP nummer
- Lexical Database for French (Base lexicale du français - BLF) : <http://ilt.kuleuven.be/blf/>
- FreeBank (geannoteerde data) <http://freebank.loria.fr/>
- Gallica <http://gallica.bnf.fr/>
- Project Gutenberg (electronische boeken) <http://www.gutenberg.org/>
- Concordancer van corpus Le Monde : <http://www.bultreebank.org/french/form.htm>

[voor een snelle zoekopdracht – data moeilijk weg te schrijven – geen precieze referenties]

- De volgende krantencorpora zijn beschikbaar op CD-rom op het secretariaat van Frans.

Le Monde 97/98	Libération 1995-1999	Le Soir 1994
Le Monde 2003-2004	Libération 2000-2001	Le Soir 1995
Le Monde okt. 2003- sept. 2005	Libération 2002-2003	Le Soir 1996
Le Monde okt. 2004- sept. 2006	Libération 2004-2005	
	Libération 2005-2006	
	Libération apr. 2005- maart 2006	

## 2.1.4 Italiaans

### 2.1.4.1 Diachrone corpora

- **TLIO** (Tesoro della Lingua Italiana delle origini)
  - Periode: teksten voorafgaand aan 1375 (proza en poëzie)
  - Grootte: 1.780 teksten /ca. 20 mil. woorden
  - Teksttypes: Gebaseerd op de OVI (Opera del Vocabolario Italiano) corpus
  - Nuttig:
  - Door:
  - Informatielink: <http://tlio.ovi.cnr.it/TLIO/index2.html>
  - Toegang:

### 2.1.4.2 Synchrone corpora

### 2.1.4.3 Gesproken corpora

- **Archivio delle Varietà dell'Italiano Parlato) / API (Archivio dell'Italiano Parlato (AVIP))<sup>11</sup>**
  - Periode:
  - Grootte:
  - Teksttypes: Conversatie analyse; Fonetische analyse (segmentaal en suprasegmentaal); task-oriented dialogen; informeel; regionaal (4 variëteiten); dialogen met (dove) kinderen
  - Nuttig:
  - Door: Department of Linguistic and Literary Studies of the University of Salerno
  - Informatielink: [www.parlaritaliano.it](http://www.parlaritaliano.it)
  - Toegang: ook beschikbaar bij prof. Crocco of in de bibliotheek (DVD)
- **Banca dati dell'italiano parlato (BADIP):** <http://badip.uni-graz.at/>
- **Corpora e Lessici di Italiano Parlato e Scritto (CLIPS)**
  - Periode: 2006
  - Grootte: 100 uren
  - Teksttypes: Conversatie analyse; Fonetische analyse (segmentaal en suprasegmentaal); task-oriented dialogen; informeel; regionaal (15 variëteiten); task-oriented dialogen; RTV (interviews, reclame)
  - Nuttig: CLIPS bevat ook een ruim corpus van schriftelijke taal:  
<http://www.ilc.cnr.it/viewpage.php/sez=ricerca/id=41/vers=ita>
  - Door:
  - Informatielink: [www.clips.unina.it](http://www.clips.unina.it)

---

<sup>11</sup> AVIP is een piloot project die uitgebreid werd met het API-project

- Toegang: Technical reports vrij, met login
  - opnames: \*.wav
  - transcripties: \*.txt

➤ **Corpus PraTiD nelle lingue europee** (PraTiD in de Europese talen)

- Periode:
- Grootte:
- Teksttypes: Conversatie analyse; Fonetische analyse (segmentaal en suprasegmentaal); 2/4 task-oriented dialogen per taal
- Nuttig: Contrastieve taalkunde (ENG, DUI, SP, PORT, FR)
- Door:
- Informatielink:
- Toegang: vrij
  - opnames: \*.wav
  - transcripties: \*.txt
  - pragmatische codering (met PraTiD): \*.xml .

➤ **Lessico di frequenza dell'Italiano Parlato (LIP)** Originele versie, boek en floppy disk met de transcripties: De Mauro T., Mancini F., Vedovelli M., Voghera M., 1993, Lessico di Frequenza dell'Italiano Parlato, Roma, Etas libri. Digitale versie: **VoLIP** (Project voor de digitalisatie van het LIP corpus; realease voorzien voor 2012)

- Periode:
- Grootte: 1.000.000 woorden
- Teksttypes: Dialogen en monologen; verschillende registers; regionaal (4 variëteiten); face-to-face dialogen; telefonische dialogen
- Nuttig: Syntaxis en tekstualiteit van de gesproken taal
- Door:
- Informatielink: <http://www.parlaritaliano.it/~parole/index.php/en/homepage>
- Toegang:

➤ **Penelope**

- Periode:
- Grootte: 30.000 woorden
- Teksttypes: Kleine stalen van verschillende variëteiten Italiaans; diachroon; oraal/geschreven; literair/ niet literair
- Nuttig: Syntaxis; Variationele analyse
- Door:
- Informatielink: <http://www.parlaritaliano.it/~parole/index.php/en/corpora/643-corpus-penelope>
- Toegang: via IBL

➤ **Corpus Stammerjohann e Corpus per il Confronto Diacronico LABLITA**

- Periode: 1965
- Grootte:
- Teksttypes: Stalen van gesproken Italiaans (uit Firenze, anno 1965); verschillende registers; Dialogen en monologen; Radiofonisch, telefonisch en face-to-face 41u15:34 (corpus Stammerjohann)
- Nuttig: POStagging (CSt); Prosodische transcripties (CSt)
- Door:
- Informatielink: <http://lablita.dit.unifi.it/corpora/confronto/>  
<http://www.parlaritaliano.it/~parole/index.php/en/corpora/646-corpus-bilanciato-consultabile-lablita>
- Toegang: vrij / op aanvraag van een gratis licentie

**2.1.4.3.1 Geschreven corpora**

➤ **CoLFIS**

- Periode: 1992-1994
- Grootte: ca. 3mil. lexicale items
- Teksttypes: Items uit kranten (artikels uit La Repubblica, La Stampa, Il Corriere della Sera) ; Artikels uit tijdschriften; Boeken
- Nuttig: Frequentie lexicon van het schriftelijke Italiaans
- Door:
- Informatielink: <http://www.ge.ilc.cnr.it/strumenti.php> en info: <http://www.istc.cnr.it/material/database/colfis/files/jadt95.pdf>
- Toegang: Vrij

➤ **CORIS/CODIS**

- Periode:
- Grootte: 120 mil. woorden
- Teksttypes: Pers; Fictie; Academische teksten; Juridische en administratieve teksten; Miscellanea; Ephemera
- Nuttig:
- Door:
- Informatielink: [http://dslo.unibo.it/coris\\_ita.html](http://dslo.unibo.it/coris_ita.html)  
info:[http://dslo.unibo.it/coris\\_ita.html](http://dslo.unibo.it/coris_ita.html)  
demo:<http://dslo.unibo.it/DEMOCORISCorpQuery.html>
- Toegang: vrij, op aanvraag

➤ **La Repubblica**

- Periode: 1985-2000

- Grootte: 400 mil. woorden
- Teksttypes: krantenmateriaal
- Nuttig: Het corpus is tokenized, POS-tagged, gelemmatiseerd, gecategoriseerd (voor genre en topic van het artikel) en opvraagbaar.
- Door:
- Informatielink: <http://dev.sslmit.unibo.it/corpora/corpus.php?path=&name=Repubblica> en info: [http://dev.sslmit.unibo.it/corpora/downloads/rep\\_lrec\\_2004.pdf](http://dev.sslmit.unibo.it/corpora/downloads/rep_lrec_2004.pdf)
- Toegang: Vrij, met login en wachtwoord

## 2.1.5 Klassieke talen

### 2.1.5.1 Latijn

#### ➤ Opera Latina/Hyperbase-Latin

- Taal: (Klassiek) Latijn
- Periode: ca. 225 v.Chr. - ca. 120 n. Chr.
- Grootte: 1.630.825 woorden
- Teksttypes: Proza en poëzie, verschillende genres. Een volledige lijst met auteurs/teksten die in het corpus zijn opgenomen kan geraadpleegd worden op <http://www.cipl.ulg.ac.be/Lasla/tlatins.html>.
- Nuttig: Dit is de enige (filologisch) geannoteerde databank die beschikbaar is; de annotatie is morfologisch/morfosyntactisch van aard (volledige morfologische determinatie van verbale en nominale vormen), maar bevat geen informatie over syntactische structuren (constituenten, clause boundaries,...)
- Door: LASLA ('Laboratoire d'Analyse Statistique des Langues Anciennes', Université de Liège) & BCL ('Base, Corpus, Langage', Université de Nice, Sophia Antipolis)
- Informatielink: <http://www.cipl.ulg.ac.be/Lasla/>
- Toegang: - De database is beschikbaar op CD-ROM (zie <http://www.unice.fr/bcl/spip.php?rubrique52&lang=nl>). Geïnteresseerden kunnen voor het raadplegen van deze CD-ROM contact opnemen met Lieven Danckaert ([Lieven.Danckaert@UGent.be](mailto:Lieven.Danckaert@UGent.be)); de CD-ROM werkt enkel onder Windows XP (misschien ook onder Vista), jammer Genoeg niet onder Windows 7.  
- De databank is ook online beschikbaar, maar hiervoor zijn gebruikersnaam en paswoord vereist (<http://www.cipl.ulg.ac.be/Lasla/login.html>).

Verder heb je met je UGent account ook toegang tot de uitgebreide site van Brepolis (<http://apps.brepolis.net/BrepolisPortal/>). Series A betreft klassiek Latijn, Series B Middeleeuws Latijn. Toegang: <http://lib.ugent.be/> > Databanken > Databanken per discipline > L&W: [LLTA : Library of Latin Texts -Series A](#).en [LLTB : Library of Latin Texts -Series B](#). Verder ook de [Vetus Latina Database](#).

#### 2.1.5.1.1 Grieks

##### ➤ Thesaurus Linguae Graecae (TLG)

- Periode: Archaïsch tot Byzantijns Grieks (VIII v.Chr. – XV n.Chr.)
- Grootte:
- Teksttypes: proza en poëzie (continu bijgewerkt)
- Nuttig: lemmatisering; aantal woorden per tekst
- Door: University of California
- Informatielink: <http://stephanus.tlg.uci.edu> (via de universitaire bibliotheek)
- Toegang: Cd-rom te ontlennen via bibliotheek Griekse Letterkunde (?)

##### ➤ Duke Databank of Documentary Papyri (DDBDP)

- Periode: Vroeg Post-klassiek tot vroeg Byzantijns Grieks (III v.Chr. – VIII n.Chr.)
- Grootte:
- Teksttypes: documentaire papyri (continu bijgewerkt)
- Nuttig: lemmatisering; met kritisch apparaat
- Door: Duke University
- Informatielink: <http://www.papyri.info> (vrij beschikbaar)
- Toegang: Cd-rom te ontlennen via bibliotheek Griekse Letterkunde (?)

#### 2.1.6 Nederlands

##### 2.1.6.1 Diachrone corpora

##### ➤ CD-ROM Middelnederlands

- Periode: 1200-1500
- Grootte:
- Teksttypes: literair proza, gedichten, ambtelijke documenten, kronieken
- Nuttig: lemmatisering, woordsoortinformatie (POS tagging)
- Door: INL
- Informatielink:
- Toegang: Cd-rom is te ontlennen via de Bibliotheek Nederlandse Taalkunde

##### ➤ Compilatiecorpus Historisch Nederlands (CHN)



- Periode: 1250-2000
- Grootte: ca.600 000 woorden
- Teksttypes: Het corpus brengt historische teksten uit verschillende bestaande bronnenverzamelingen samen tot
- Nuttig: de compilatie is nuttig omdat het een methodologisch verantwoorde steekproef is die de gehele geschiedenis van het Nederlands overspant.
- Door: Evie Coussé
- Informatielink: <http://www.flw.ugent.be/eviecousse/corpus>
- Toegang: vrij

Meer info en links naar diachrone corpora op de website van het Instituut voor Nederlandse Lexicologie (<http://www.inl.nl>), i.h.b. <http://www.inl.nl/tst-centrale/nl/producten/corpora/6> :

- Corpus Gysseling Corpus Oudnederlands
- Delftse bijbel 1477 Deux-Aesbijbel 1562
- Leuvense bijbel 1548
- Lutherse bijbel 1648
- Statenvertaling 1637

## **2.1.6.2 Synchronie corpora**

### **2.1.6.2.1 Gesproken corpora**

- **Corpus Gesproken Nederlands (CGN)**
  - Periode: 1998-2004
  - Grootte:
  - Teksttypes: gesproken conversatie
  - Nuttig: lemmatisering, woordsoortinformatie (POS tagging), zinsgrenzen, boomstructuren
  - Door: Nederlandse Taalunie
  - Informatielink: <http://lands.let.kun.nl/cgn/> en [http://www.tst.inl.nl/cgndocs/doc\\_Dutch/topics/index.htm](http://www.tst.inl.nl/cgndocs/doc_Dutch/topics/index.htm)
  - Toegang: Er is een cd-rom beschikbaar via de Bibliotheek Nederlandse Taalkunde

Meer info en links naar gesproken corpora op de website van het Instituut voor Nederlandse Lexicologie (<http://www.inl.nl>), i.h.b. <http://www.inl.nl/tst-centrale/nl/producten/corpora/6> :

- CHOREC-spraakcorpus
- Corpus Pathologische en Normale Spraak (COPAS)
- D-TUNA-corpus
- IFA-corpus
- IFA-dialoog-videocorpus
- JASMIN-spraakcorpus

### 2.1.6.2.2 Geschreven corpora

Info en links naar geschreven corpora op de website van het Instituut voor Nederlandse Lexicologie (<http://www.inl.nl>), i.h.b. <http://www.inl.nl/tst-centrale/nl/producten/corpora/6>:

- 27 Miljoen Woorden Krantencorpus 1995
- 38 Miljoen Woorden Corpus
- 5 Miljoen Woorden Corpus 1994
- AUTONOMATA-namencorpus
- AUTONOMATA-POI-corpus
- COREA-coreferentiecorpus
- DAESO-corpus: Parallele Nederlandstalige monolinguale treebank
- DPC-corpus: Dutch Parallel Corpus
- Eindhoven Corpus
- Lassy Groot-corpus
- Lassy Klein-corpus
- PAROLE Distributable Corpus
- PAROLE-corpus 2004
- SoNaR
- SumNL-samenvattingencorpus

### 2.1.6.3 Literatuur

- In het digitale tijdschrift DigiTaal (<http://www.niederlandistik.fu-berlin.de/digitaal>) vind je enkele artikels waarin Nederlandstalige corpora besproken worden.
- <http://lib.ugent.be/> > Databanken > Databanken per discipline > L&W: [ANS](#) : [Algemene Nederlandse Spraakkunst](#).
- <http://lib.ugent.be/> > Databanken > Databanken per discipline > L&W: [EWN](#) : [Etymologisch Woordenboek van het Nederlands](#).

## 2.1.7 Portugees

### 2.1.7.1 Diachrone corpora

- **Corpus do Português**
  - Periode: 1300-2000
  - Grootte: 45 miljoen woorden
  - Teksttypes: opdeling in Academisch (hoofdzakelijk Encarta), gesproken, pers en fictie -
  - Nuttig: zowel Braziliaanse Portugese als Continentaal Portugese teksten –

mogelijkheid tot opzoeken woordklassen - nadelen: geen toegang tot de volledige teksten; enkel opzoekingen mogelijk voor een volledige eeuw (geen mogelijkheid de opzoekingen te beperken tot bv. voor en na 1950)

- Door: Michael Ferreira (Georgetown University) & Mark Davies (Brigham Young University)
- Informatielink: <http://corpusdoportugues.org/>
- Toegang: vrij

#### ➤ **Corpus Histórico do Português**

<http://www.tycho.iel.unicamp.br/~tycho/corpus/index.html>

**2.1.7.2** Synchrona corpora

**2.1.7.3** Gesproken corpora

**2.1.7.4** Geschreven corpora

- **Corpus do Português** (cf. supra)
- **O Constructor** : <http://www.geocities.com/Athens/Crete/1546/indexc.htm>
- **Corpus PALOP de literatura poscolonial portugués-español**:  
<http://sli.uvigo.es/CLUVI/>
- **Corpora disponibilizados pelo CLUL da Universidade de Lisboa**:  
[http://www.clul.ul.pt/sectores/projecto\\_crpc.html#disponibiliz](http://www.clul.ul.pt/sectores/projecto_crpc.html#disponibiliz)
- **Linguística recursos digitais (Centro virtual Camões)**:  
<http://www.instituto-camoes.pt/cvc/reclinguistica/linguistica.html>
- **Linguateca**: <http://www.linguateca.pt>

### **2.1.8 Scandinavistiek**

**2.1.8.1** Diachrone corpora

- **Fornsvenska textbanken** (Zweeds):  
<http://www.nordlund.lu.se/Fornsvenska/Fsv%20Folder/>

**2.1.8.2** Synchrona corpora

**2.1.8.2.1** Gesproken corpora

- **Göteborg Spoken Language Corpus (GSLC)** (Zweeds):  
<http://www.ling.gu.se/projekt/tal/>

**2.1.8.2.2** Geschreven corpora

- **Språkbanken**
  - Taal: Zweeds

- Periode:
- Grootte: 150 à 200 miljoen woorden
- Teksttypes: fictie en non-fictie
- Nuttig:
- Door:
- Informatielink: <http://spraakbanken.gu.se>
- Toegang:

➤ **PAROLE-korpusen**

- Taal: Zweeds
- Periode:
- Grootte: 19,4 miljoen woorden
- Teksttypes: fictie en non-fictie
- Nuttig: geannoteerd corpus
- Door:
- Informatielink: <http://spraakbanken.gu.se/parole/>
- Toegang:

➤ **Stockholm Umeå Corpus (SUC)**

- Taal: Zweeds
- Periode:
- Grootte: 1 miljoen woorden
- Teksttypes: gemengde genres; korte teksten (max. 2000 woorden/tekst); krantenartikels
- Nuttig: geannoteerd corpus
- Door:
- Informatielink:
- Toegang: vrij toegankelijk via Parole

➤ **KorpusDK (Deens):** <http://ordnet.dk/korpusdk>

➤ **SALT Dut-Swe korpus**

- Taal: Nederlands - Zweeds
- Periode:
- Grootte: 3 miljoen woorden
- Teksttypes: parallel vertaalcorpus
- Nuttig:
- Door: Gudrun Rawoens
- Informatielink: cf. Rawoens (2010)
- Toegang: nog niet online toegankelijk

- **Finlandssvensk textkorpus: Språkbanken i Finland** (Finlandszweeds):

<http://www.kotus.fi/index.phtml?l=sv&s=2075>

- **Litteraturlbanken** (Zweeds, fictie):

<http://litteraturlbanken.se>

- **Project Runeberg** (Zweeds, oudere fictie):

<http://runeberg.org/>

### 2.1.8.3      **Andere bronnen (o.a. woordenboeken)**

- **Svenska Akademiens ordbok (SAOB)**

- historisch woordenboek Zweeds

- <http://www.saob.se>

- **Svenska Akademiens ordlista (SAOL)**

- woordenlijst Zweeds

- [www.svenska.akademien.se/web/Ordlista](http://www.svenska.akademien.se/web/Ordlista)

- **Söderwall och Schlyter**

- woordenboek oud-Zweeds

- toegankelijk via Språkbanken

- **Svensk Etymologisk Ordbok**

- etymologisch woordenboek Zweeds

- <http://runeberg.org/svetym/>

### 2.1.8.4      **Literatuur**

Delsing, Lars-Olof (2002). Fornsvenska textbanken. I: Lagman, Svante, stig Örjan Ohlsson, Viivika Voodla (eds), *Svenska språkets historia i Östersjöområdet*. Tartu, s. 149-156.

Hellquist, Elof (1957). *Svensk Etymologisk Ordbok*. Tredje upplagan. Band I: A-N. Lund: Gleerups Förlag. Available via Project Runeberg <<http://runeberg.org/>>

Holmbäck, Åke & Elias Wessén (1946). *Svenska landskapslagar : tolkade och förklarade för nutidens svenskar*. Stockholm: Hugo Gebers Förlag.

SAOB = *Ordbok över Svenska Språket utgiven av Svenska Akademien*. (1898–). Lund: Gleerups. (volume 16 published in 1942).

Rawoens, Gudrun (2010). Multilingual corpora in cross-linguistic research. Focus on the compilation of a Dutch-Swedish parallel corpus. I: Sergio Bolasco, Isabella Chiari and Luca Giuliano (eds), *Statistical Analysis of Textual Data. Proceedings of 10th International Conference JADT2010*, LED Edizioni Universitarie, Milano, s. 1287–1294.

Schlyter, C.J. (1877). *Ordbok till Samlingen af Sweriges Gamla Lagar*. (Saml. af Sweriges Gamla Lagar 13.) Lund. Available via Fornsvensk Lexikalisk Databas on <<http://spraakbanken.gu.se/fsvldb>> (last accessed AprXXil 7, 2010)

- Söderwall, K.F. (1884-1918). *Ordbok Öfver svenska medeltids-språket*. Vol I-III. Lund. [22.894 uppslagsord]. Available via Fornsvensk Lexikalisk Databas on <http://spraakbanken.gu.se/fsvldb>
- Söderwall, K.F. (1953-1973). *Ordbok Öfver svenska medeltids-språket*. Supplement. Vol IV—V. Lund. [21.495 uppslagsord]. Available via Fornsvensk Lexikalisk Databas on <http://spraakbanken.gu.se/fsvldb>

## 2.1.9 Spaans

### 2.1.9.1 Diachrone corpora

#### ➤ Corpus del Español (CDE)

- Periode: 1200-2000
- Grootte: 100 miljoen woorden
- Teksttypes: zowel Hispano-Amerikaanse als Spaanse teksten - mogelijkheid tot opzoeken woordklassen - opdeling in Academisch (hoofdzakelijk Encarta), gesproken, pers en fictie - nadelen: geen toegang tot de volledige teksten; geen mogelijkheid zoekopdrachten te beperken tot een bepaalde geografische regio; enkel opzoeken voor een volledige eeuw (geen mogelijkheid de opzoeken te beperken tot bv. voor en na 1950)
- Nuttig: woordsoortinformatie (POS tagging)
- Door: Mark Davies, Brigham Young University
- Informatielink: <http://corpusdelespanol.org/>
- Toegang: vrij

#### ➤ Corpus Diacrónico del Español (CORDE)

- Periode: begin-1975
- Grootte: ca. 23 miljoen woorden
- Teksttypes: historisch corpus vanaf de eerste Spaanse teksten tot 1975 - diverse registers: zowel geschreven als gesproken (voor de 20e eeuw)
- Nuttig:
- Door: Real Academia Española
- Informatielink: <http://corpus.rae.es/cordenet.html>
- Toegang:

### 2.1.9.2 Synchrone corpora

#### 2.1.9.2.1 Gesproken corpora

#### ➤ Macrocorpus de la norma lingüística culta de las principales ciudades del mundo hispánico

- Periode: 1960-2000
- Grootte:

- **Teksttypes:** Het corpus omvat gestuurde interviews van een half uur (gemaakt tussen de jaren 60 en 90) met burens uit 12 grote steden in Spanje en Latijns-Amerika (14 interviews in elke stad, gesorteerd op basis van geslacht en leeftijd van de ondervraagden).
  - **Nuttig:** De CD-ROM bevat geen opnamen, of transcripties met fonetische symbolen, maar traditionele transcripties die nuttig zijn voor morfosyntactische, lexicale, semantische en misschien discursieve analyse.
  - **Door:**
  - **Informatielink:**
  - **Toegang:** Cd-Rom Spaanse Taalkunde
- **Corpus OralRom** (Bib. Franse taalkunde, L 14B)
- **Periode:**
  - **Grootte:** ca. 300.000 woorden/ taal
  - **Teksttypes:** Het omvat opnames, fonetische diagrammen en transcripties gesegmenteerd in uitingen voor het Frans, Spaans, Italiaans en Portugees
  - **Nuttig:** Spontane gesprekken die kunnen worden geanalyseerd op elk taalkundig niveau, met inbegrip van het fonetische. Nadeel is dat het de transfert naar MS Office-documenten niet toelaat.
  - **Door:**
  - **Informatielink:**
  - **Toegang:** Bib. Franse tlk. (L 14B)
- **Corpus COLAm** (taal van Madrileense jongeren) aanvragen via <http://gandalf.aksis.uib.no/cola/adgang/index-s.html>

#### **2.1.9.2.2 Geschreven corpora**

- **Corpus de Datos Sintácticos del Español Actual (BDS):** <http://www.bds.usc.es/>
- **Corpus del Español** (cf. supra)
- **Corpus de Referencia del Español Actual (CREA)**
- **Periode:** 1975-2004
  - **Grootte:** ca. 154 miljoen woorden
  - **Teksttypes:** literair proza, theaterteksten, krantenmateriaal
  - **Nuttig:** 50% Spaans uit Spanje; 50% Spaans uit Latijns-Amerika - zowel geschreven

(90% boeken, pers en varia) als gesproken (10%) registers - thematische opdeling - nadeel: geen toegang tot volledige teksten, enkel toegang via de queries

- Door: Real Academia Española
- Informatielink: <http://corpus.rae.es/creanet.html>
- Toegang: vrij

➤ **Logos-Wordtheque:** <http://www.wordtheque.com>

## 2.1.10 Het internet als corpus

Elke verzameling teksten vormt in principe een corpus. In die zin is het internet op zich ook een corpus waar je met behulp van een zoekmachine (Google, altavista, etc.) een oppervlakkig taalkundig onderzoek kan uitvoeren (Sanders 2011). Meer precieze opdrachten kunnen via de volgende tools uitgevoerd worden.

### ➤ **Glossanet II**

- Periode: hedendaags
- Grootte: -
- Teksttypes: Zoekmotor voor krantenteksten op het internet. Franse of Engelse interface. Verschillende talen: Engels, Fins, Frans, Duits, Grieks, Italiaans, Koreaans, Noors, Pools, Portugees, Braziliaans, Russisch, Servisch, Spaans, Thai, Nederlands.
- Nuttig: dagelijkse search van je geformuleerde zoekopdracht in de aangeduide kranten. Je ontvangt het materiaal in je mailbox.
- Door: Université Catholique de Louvain
- Informatielink: <http://glossa.fltr.ucl.ac.be/>
- Toegang: vrij na registratie

### ➤ **Webcorp Live**

- Periode: real time
- Grootte: -
- Teksttypes: Zoekmotor voor internetsites allerhande
- Nuttig: voorziet enhanced sentence, boundary detection, date identification, collocation and other statistical analyses, grammatical tagging, language detection, full pattern matching and wildcard search.
- Door: Research and Development Unit for English Studies (RDUES) in the School of English at Birmingham City University.
- Informatielink: <http://www.webcorp.org.uk/>
- Toegang: vrij



### 2.1.11 Parallele corpora

➤ **Dutch Parallel Corpus (DPC):**

<http://www.kuleuven-kortrijk.be/dpc/manual/Introduction.html>

➤ **Linguee :** <http://www.linguee.fr/francais-anglais/page/about.php>

➤ **Namur Corpus**

- Periode: 1982-1993
- Grootte: 2,000,000 woorden
- Genre: Parallel vertaalcorpus (Engels, Nederlands en Frans). Het corpus bevat originele teksten (in elke taal), en hun parallelle vertaling in elk van de twee overige talen.
- Teksttypes: Fictie: Extracten uit 12 originele fictieteksten (< boeken gepubliceerd tussen 1982 en 1992). Non-fictie: (a) extracten uit de UNESCO Courier (b) transcripties van debatten uit het Europees Parlement
- Nuttig: voor contrastief onderzoek in Engels, Frans en / of Nederlands. Eender welke paragraaf die in één van de drie talen geselecteerd wordt, kan automatisch en makkelijk vergeleken worden met de corresponderende paragraaf in de twee andere talen.
- Nadeel: Opzoeken kunnen enkel gedaan worden via Hans Paulussen. Je moet dus heel specifiek doorgeven wat je wil (laten) opzoeken. De snelheid waarmee je resultaten opgestuurd krijgt, hangt dus ook soms af van de tijd die hij er voor kan vrijmaken...
- Door: Hans Paulussen – Universiteit van Namen; (Universiteit Gent)
- Informatielink: <http://www.kuleuven-kortrijk.be/~hpauluss/NamurCorpus.html>  
[http://www.kuleuven-kortrijk.be/~hpauluss/NC/NC\\_descr.html](http://www.kuleuven-kortrijk.be/~hpauluss/NC/NC_descr.html)
- Beschikbaarheid: Via Hans Paulussen die de gevraagde opzoeken zelf doet en doorstuurt.

➤ **Open Parallel corpus (Opus)** <http://opus.lingfil.uu.se/> met links naar andere parallelle corpora (bv. Belgisch Staatsblad, Europarl, etc.)

➤ **The JRC-Acquis Multilingual Parallel Corpus** (<http://langtech.jrc.it/JRC-Acquis.html>)

### 2.1.12 Learner corpora

- Corpus de dissertations françaises (**CODIF**) en Label France French to English bilingual corpus <sup>12</sup>
- French Interlanguage Database (**Frida**) <http://www.fltr.ucl.ac.be/fltr/germ/etan/cecl/Cecl-Projects/Frida/fridatext.htm>
- Louvain International Database of Spoken English Interlanguage (**LINDSEI**), vervolg op International Corpus of Learner English (**ICLE**): <http://www.uclouvain.be/en-cecl-icle.html>
- **Interfra**: <http://www.fraitaklass.su.se/english/interfra>
- <http://talkbank.org/>
- **FLLOC**: <http://www.flloc.soton.ac.uk/>
- **SPLLOC**: <http://www.splloc.soton.ac.uk/>
- **LIPS** (Lessico Italiano Parlato di Stranieri) IT-L2
  - Periode:
  - Grootte: 100 uren gesproken taal; 700.000 woorden
  - Teksttypes: transcripties  
Frequentielijsten; 2000 examen CILS (Certificazione di Italiano come Lingua Straniera); dialogen en monologen van leerlingen met verschillende L1
  - Nuttig:
  - Door:
  - Informatielink: <http://www.parlaritaliano.it/~parole/index.php/en/corpora/653-corpus-lips>
  - Toegang: vrij

## 2.2 Gegevens selecteren: Het gebruik van concordancers

Een **concordancer** (of concordantiesoftware<sup>13</sup>) wordt binnen de corpustaalkunde gebruikt om een gesorteerde lijst met taalkundige data uit een corpus te extraheren om daarna verder door de corpustaalkundige geanalyseerd te worden, bijvoorbeeld in een databank als Access (§2.3). Bepaalde corpora zijn al geannoteerd (bv. gelemmatiseerd, POS tagged) en beschikken over een geïntegreerde concordancer (bv. Frantext, BNC). Dit bespaart u enorm veel tijd. Is de

---

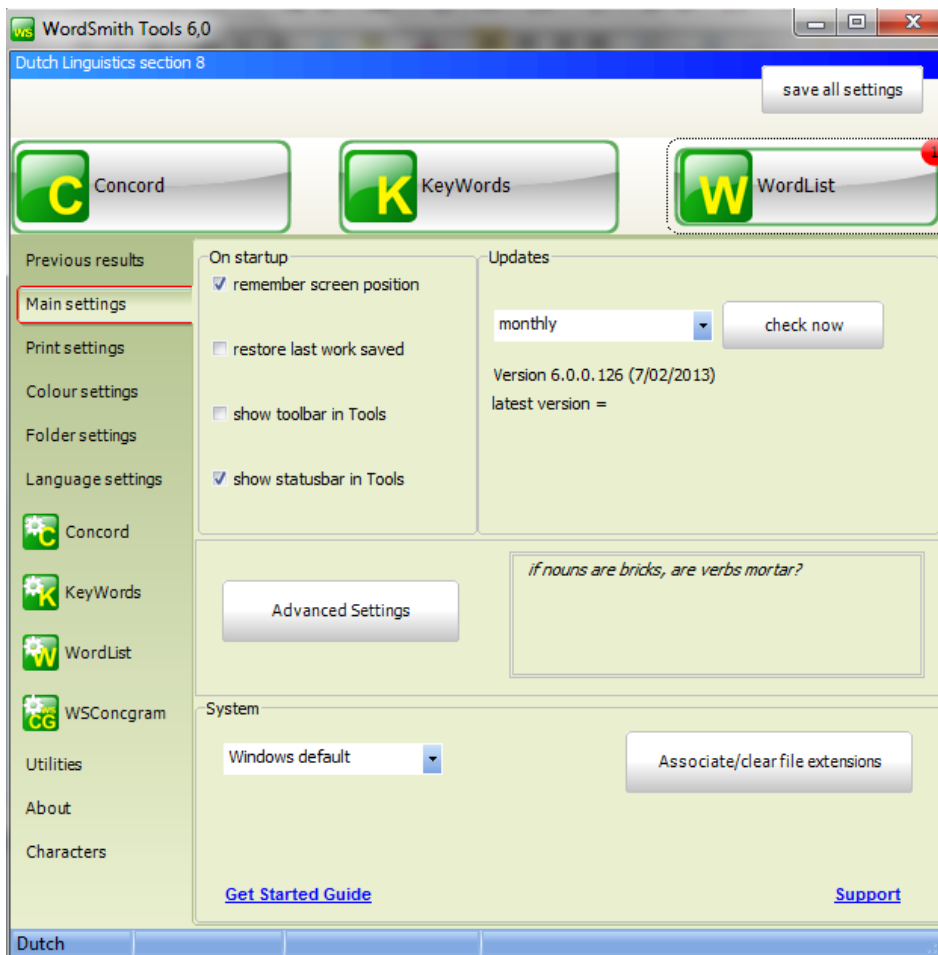
<sup>12</sup> Enkele belangrijke *learner corpora* voor de studie van taalverwerving zijn gecreëerd door het Centre for English Corpus Linguistics aan de UCL. Een overzicht van hun corpora vindt u hier: <http://www.uclouvain.be/en-258636.html>.

<sup>13</sup> Voor een lijst van concordantiesoftware en andere interessante tools zie o.a. <http://linguistlist.org/sp/GetWRListings.cfm?WRAbbrev=Software>

selectie niet zo makkelijk, dan moet u aan de slag met een concordancer die u loslaat op, bijvoorbeeld, een krantencorpus. Binnen de vakgroep Taalkunde aan de Universiteit Gent is vooral expertise opgebouwd met betrekking tot de concordancer WordSmith Tools. De versie WordSmith Tools 6.0 is geïnstalleerd (sinds maart 2013) op alle computers in PC lokaal A en B van de Blandijn. Daar kan je het programma starten door te klikken op de gelijknamige snelkoppeling in de map P:\LW. In wat volgt bespreken we de woordenlijst en concordfunctie. We gaan niet dieper in op de sleutelwoordfunctie.

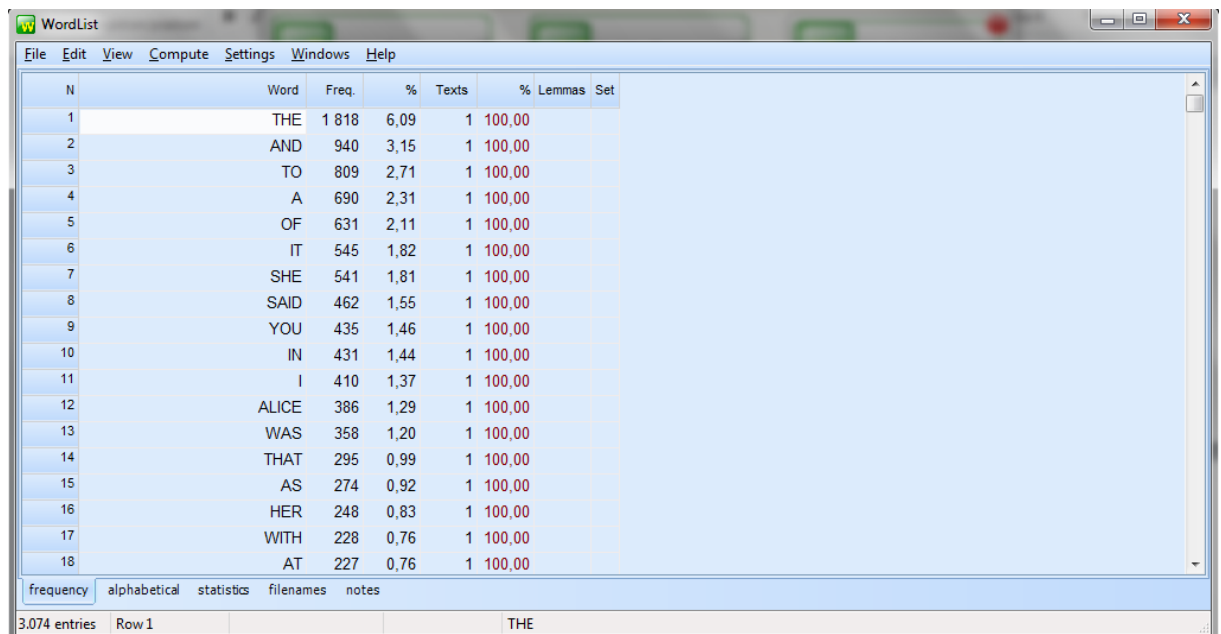
## 2.2.1 WordSmith Tools 6.0: WordList functie

1. Open WordSmith. Het hoofdmenu verschijnt als volgt.



2. Stel de 'Language setting' in voor uw corpus.
3. Voor lexicale distributie, klik op WordList – File – New
4. In het tabblad 'Main' ga je naar 'Choose texts now'. Je selecteert een tekst die je zonet hebt opgeslagen en die bij voorkeur een Plain text format (.txt extensie) heeft. Sleep het naar 'Files selected'. Klik op dat bestand en dan op OK.

5. Voor het document ‘Alice in Wonderland’ (Project Gutenberg) krijgen we de volgende resultaten:



The screenshot shows the WordList application window with a menu bar (File, Edit, View, Compute, Settings, Windows, Help) and a table of word frequencies. The table has columns: N, Word, Freq., %, Texts, % Lemmas, and Set. The data is sorted by frequency, with 'THE' being the most frequent word.

N	Word	Freq.	%	Texts	% Lemmas	Set
1	THE	1 818	6,09	1	100,00	
2	AND	940	3,15	1	100,00	
3	TO	809	2,71	1	100,00	
4	A	690	2,31	1	100,00	
5	OF	631	2,11	1	100,00	
6	IT	545	1,82	1	100,00	
7	SHE	541	1,81	1	100,00	
8	SAID	462	1,55	1	100,00	
9	YOU	435	1,46	1	100,00	
10	IN	431	1,44	1	100,00	
11	I	410	1,37	1	100,00	
12	ALICE	386	1,29	1	100,00	
13	WAS	358	1,20	1	100,00	
14	THAT	295	0,99	1	100,00	
15	AS	274	0,92	1	100,00	
16	HER	248	0,83	1	100,00	
17	WITH	228	0,76	1	100,00	
18	AT	227	0,76	1	100,00	

At the bottom of the window, there are tabs for 'frequency', 'alphabetical', 'statistics', 'filenames', and 'notes'. The 'frequency' tab is selected. Below the tabs, it says '3.074 entries' and 'Row 1'. The word 'THE' is highlighted in the 'Word' column.

Naast een ranking volgens frequentie, kunnen we ook een alfabetische ranking opvragen (onderaan). Het tabblad ‘statistics’ biedt een kwantitatieve samenvatting van alle data.

## 2.2.2 WordSmith Tools 6.0: Concord functie

De Concord functie is de belangrijkste functie die u de nodige data verschaft voor allerhande verdere analyses (syntactisch, semantisch,...).

1. Ga naar het hoofdmenu en klik op ‘Concord – File – New’

2. In het tabblad ‘Texts’ selecteer je de tekst zoals eerder besproken.

1. In het tabblad ‘Search Word’ begeleidt volgende legende je zoekactie.

book  
(gets book but not books, booking etc.)

book\*  
(gets book OR booking OR books etc.)

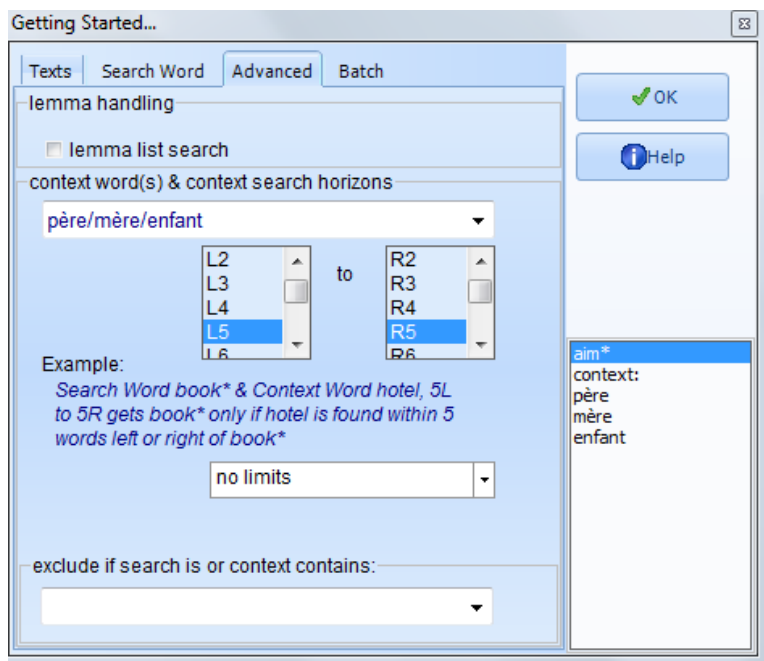
book\* a hotel  
(gets 'book a hotel' OR 'booking a hotel' but not 'book a good hotel')

==New==  
(gets New but not new)

Paris/Moscow/New York  
(gets Paris OR Moscow OR New York)

Let op!:

- de Boleaanse operator 'OR' in de legende betekent: Het resultaat is waar als "A" waar is of als "B" waar is of als beide waar zijn. Het resultaat is onwaar als zowel "A" als "B" onwaar zijn. M.a.w., als je met de asterisk werkt, bv. 'aim\*' voor alle flexies van het werkwoord *aimer*, dan zoekt WordSmith alle woorden die beginnen met 'aim'. Nadeel is dat je eventueel ook niet gewenste woorden als resultaat kunt krijgen (*aimable*, *aimanter*...).
- WordSmith is accentgevoelig: voeg ze toe waar nodig (à, ç, é, è, ï, etc.). WordSmith is niet hoofdlettergevoelig, tenzij je met = werkt (cf. legende 'New').
- In plaats van alle vormen in te voeren, gescheiden door een "/" : aime/aimes/aiment/aime kan je voor een uitgebreide zoekactie ook een document toevoegen ('search-words from a file') dat bijvoorbeeld alle flexies van het lemma bevat.
- Ga naar het tabblad 'Advanced'. Als je naar een combinatie van woorden zoekt, typ je onder 'Context Word(s) & Context Search Horizons' het volgende woord van de combinatie. Je kunt bv. zoeken naar combinaties van "aimer" gevolgd door 'père/mère/enfant'. Je kunt hier ook nog eens aanduiden waar het tweede woord moet voorkomen : bv. binnen een bereik van 5 woorden links of 5 woorden rechts.



- Als het resultaat verschijnt (met de gezochte vorm centraal in kleur), kan je de weergave van de data bewerken in de tabbladen 'Edit' en 'View' (boven). De tabbladen 'collocates', 'patterns' en 'clusters' (onderaan) kunnen interessante statistische informatie bevatten.

N	Concordance	Set	Word #	Sen	Sen	Para	Para	Hea	Hea	Sec	Sec	File	Date	%
1	nice little dog near our house I should <b>like</b> to show you! A little bright-eyed	A	4 145	192	89%	0	14%			0	14%	alice.txt	2013/apr/17 00:	13%
2	said the Caterpillar. 'Well, I should <b>like</b> to be a LITTLE larger, sir, if you		9 577	530	19%	0	33%			0	33%	alice.txt	2013/apr/17 00:	31%
3	Turtle repeated thoughtfully. I should <b>like</b> to hear her try and repeat		21 830	1 3	33%	0	74%			0	74%	alice.txt	2013/apr/17 00:	73%
4	carrying clubs; these were all shaped <b>like</b> the three gardeners, oblong and		16 112	957	26%	0	55%			0	55%	alice.txt	2013/apr/17 00:	53%
5	because the chimneys were shaped <b>like</b> ears and the roof was thatched		13 409	750	80%	0	46%			0	46%	alice.txt	2013/apr/17 00:	44%

- Je kan nu het volledige resultaat opslaan in een bestand (File – Save as) als tekst- of rekenbladformaat. Als je het als *Plain text* opslaat (.txt format) moet je eerst definiëren hoeveel karakters je wil behouden per voorbeeld, te bepalen in 'Concordance Characters'. Weet dat 500 karakters vaak geen luxe is als je een minimale contextuele analyse wil uitvoeren.
- Word behoort niet tot de standaard mogelijkheden om op te slaan. Wil je het resultaat liever niet in Excel<sup>14</sup> maar in Word bekijken en verder bewerken, dan kan je best eerst het document opslaan als een Worddocument : ga naar 'Opslaan als' en kies onder 'opslaan als': 'word-document (.docx)'. Om het geheel overzichtelijker te maken, selecteer je het hele document, ga je naar 'Tabel', dan naar 'converteren Tekst naar tabel'. Onder 'lijstscheidingstekens' klik je op 'tab'. Het resultaat is dan normaal een kolom met voorbeelden, netjes onder elkaar en gescheiden van de andere info.

## 2.3 Gegevens beheren: Het gebruik van een database<sup>15</sup>

Met **Access** en **Filemaker Pro**<sup>16</sup> is het mogelijk om de resultaten van een zoekopdracht met bijvoorbeeld Wordsmith op een ordelijke manier op te slaan en te sorteren in een databank. Bovendien laten de programma's ook toe om de gegevens aan te vullen met taalkundige informatie, d.w.z. om er de voorbeelden verder in te analyseren. Ten slotte zijn de programma's interessant om heel gericht de databank te bevragen en verschillende onderzochte parameters met elkaar in verband te brengen.

<sup>14</sup> Het voordeel van Excel is de mogelijkheid om de analyses direct om te zetten in een draaitabel (§ 2.4).

<sup>15</sup> Red.: Annemie Demol, februari 2010.

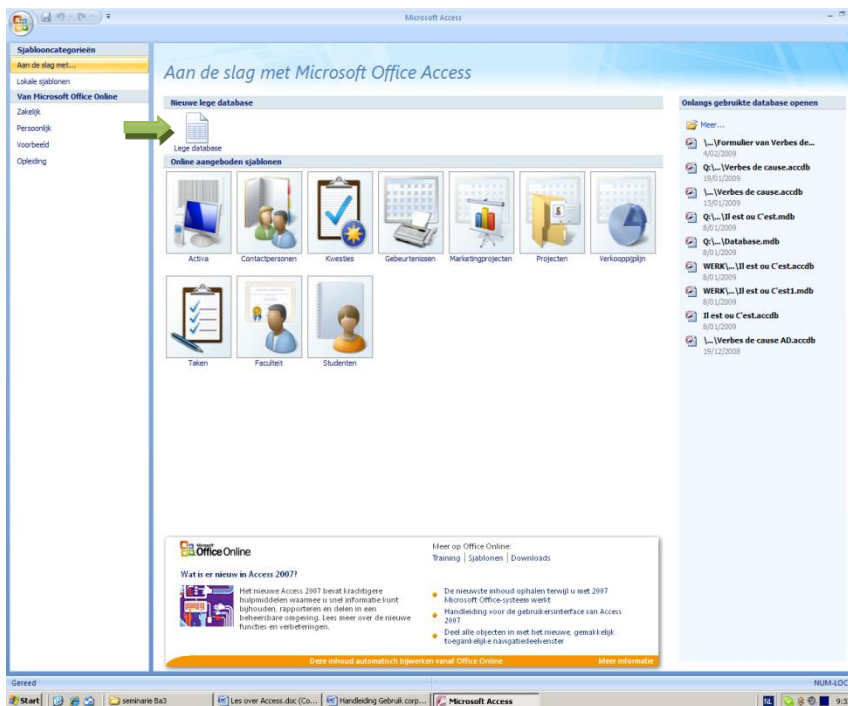
<sup>16</sup> Het nadeel van Filemaker Pro t.o.v. Access is dat het om een betalende licentie gaat. Access daarentegen maakt meestal deel uit van het courante Microsoft Office pakket. We beperken ons hier tot deze software.

### 2.3.1 Inleiding tot Access

Microsoft Office Access 2007 is een programma bestemd om databases of “gegevensbanken” mee te creëren. Het wordt voor heel wat (niet taalkundige) toepassingen aangewend én kan nuttig zijn voor wie taalkundige gegevens wil ordenen, analyseren en er bewerkingen op wil uitvoeren. Het kan dan gaan om corpusgegevens, maar ook om data verkregen uit enquêtes of testen.

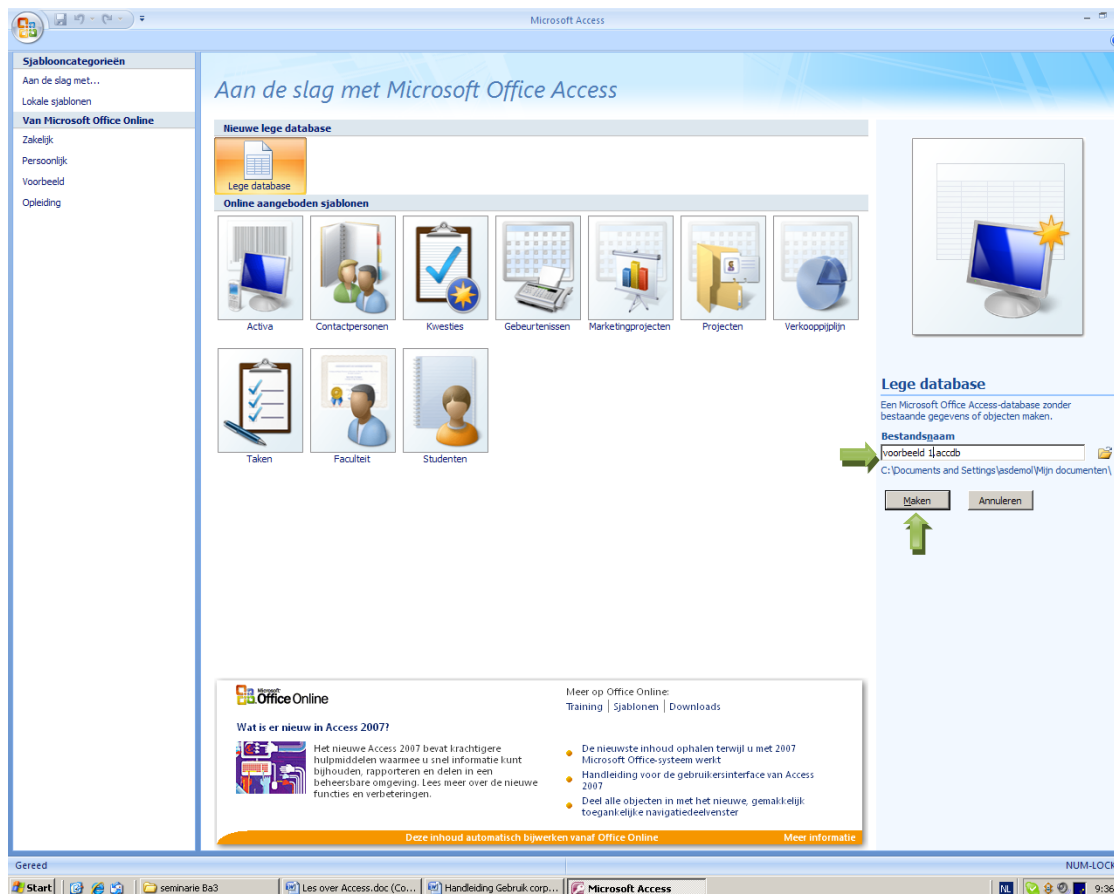
Access heeft een aantal voordelen ten opzichte van Excel/Word: het laat meer ingewikkelde bewerkingen toe en dankzij de formulieren kun je ook overzichtelijker werken. Het voorbeeld dat je analyseert, blijft toegankelijk op elk moment. Resultaten van zoekacties en berekeningen kunnen gekopieerd worden naar Excel, om er aangepaste tabellen en allerhande grafieken mee te maken. Initiële gegevens, zoals onder meer de voorbeelden die je via WordSmith hebt verkregen, kunnen vanuit Word geïmporteerd worden in Access.

Hoe start je Access op? Zoals andere MS Office toepassingen: klik op start (onderaan je bureaublad), selecteer “programma’s”, daarna “microsoft office” en dan klik je op “Microsoft Office Access”. Dit is het beginvenster:



Klik bovenaan links op “lege database” (cf. groene pijl).

In het deelvenster dat rechts verschijnt, typ je de naam die je aan je database wil geven. Behoud de extensie .accdb. Klik op het mapje om aan te geven waar je de database wil opslaan. Klik daarna op “maken”.



## 2.3.2 Tabel ontwerpen

Het venster dat zich automatisch opent, brengt je bij de tabel. In de tabel worden alle gegevens die je invoert, verzameld. Het ontwerp van de tabel is van essentieel belang: je kunt alleen opzoeken verrichten en bewerkingen uitvoeren op de gegevens die je erin stopt. Al klinkt dit nog zo evident, het kan wel voor problemen zorgen.

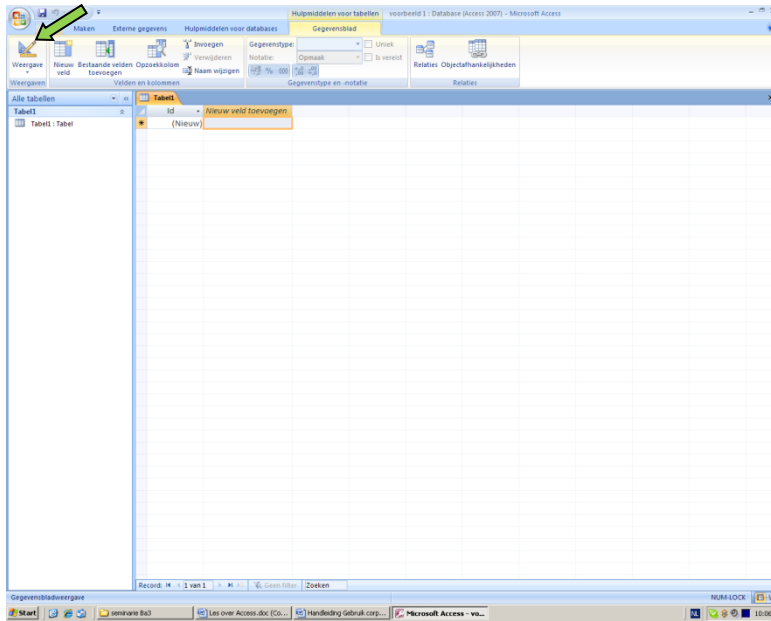
Het moeilijkste onderdeel van het werken met Access, bestaat er namelijk in om, voor je begint, heel goed na te denken over **wat** je precies wilt aanvangen met je data: wat wil je eigenlijk analyseren, welke gegevens zal je (waarschijnlijk) nodig hebben? Voor je aan de tabel begint, heb je dus een lijstje nodig met criteria die je wil onderzoeken in je voorbeelden: dat lijstje maak je op aan de hand van wat je al over je onderwerp gelezen hebt, maar ook door er eventueel een grammatica of woordenboek op na te slaan, door gewoon goed na te denken over je studieobject - over alles wat eventueel van belang zou kunnen zijn - en eventueel door eerst een aantal voorbeelden aandachtig door te lezen om te na te gaan of er je iets opvalt.<sup>17</sup>

<sup>17</sup> Als je gegevens verkregen zijn aan de hand van enquêtes of testen, dan zullen de items in je tabel natuurlijk overeenstemmen met de verschillende vragen of oefeningen die je hebt voorgelegd aan je respondenten.



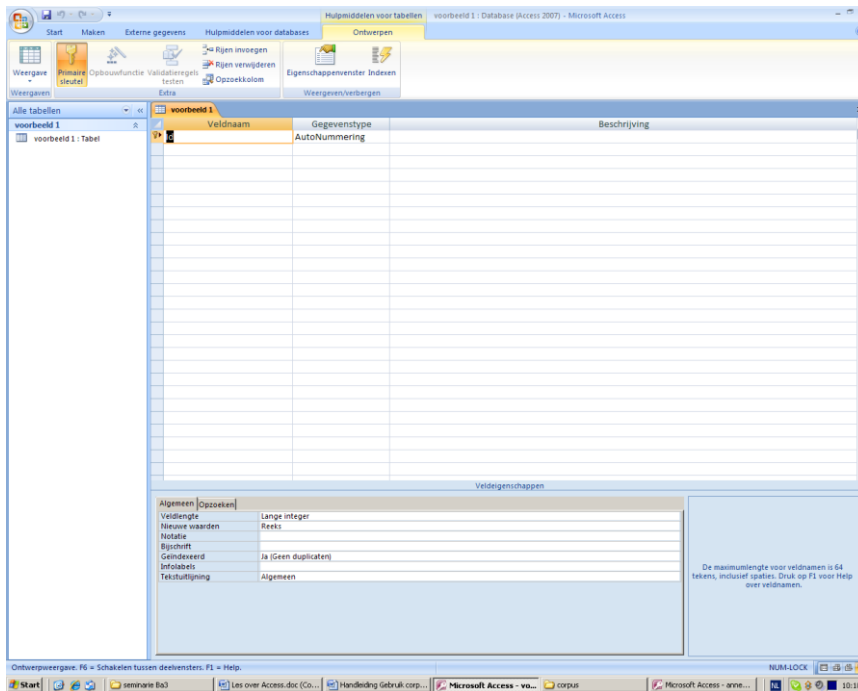
Eens je dat lijstje klaar hebt, moet er worden nagedacht over de manier waarop je de gegevens aan de tabel zult toevoegen. Voor je kunt beginnen je voorbeelden te importeren naar Access en ze te analyseren, moet je inderdaad je tabel ontwerpen: je moet je tabel eerst klaarmaken door in te geven welke criteria je zult analyseren en hoe je dat zult doen. Dit wordt hieronder stap voor stap uitgelegd en geïllustreerd met een voorbeeld.

Het venster dat je nu voor je ziet, is eigenlijk de gegevensweergave van de tabel. Door in de linkerbovenhoek te klikken op het icoon “weergave”, kun je overschakelen naar de ontwerpweergave van de tabel:



Er wordt je dan wel eerst gevraagd om je tabel een naam te geven. Die naam kies je natuurlijk zelf, in functie van je studieobject. Als je studie gaat over het gebruik van en de verschillen tussen de twee werkwoorden *aboutir à* en *déboucher sur*, dan kun je bijvoorbeeld de volgende naam geven: **verbes**.<sup>18</sup> Daarna opent zich de ontwerpweergave:

<sup>18</sup> Als je werkt met een oudere versie van Access, dan is het aan te raden om de naam van je tabel te laten voorafgaan door T, zodat je weet dat het om een tabel gaat (en niet om een formulier bv., cf. infra).



Er wordt je nu gevraagd om eerst een primaire sleutel te definiëren. Access doet dit eigenlijk al automatisch voor jou door in de eerste rij van de ontwerpweergave 'id' in te vullen en als gegevenstype autonummering te selecteren. Die primaire sleutel is nodig als je verschillende tabellen met elkaar wilt linken. Dit is al een heel complex ontwerp voor een database en is in de meeste gevallen overbodig. Bovendien vormt die primaire sleutel een probleem bij het kopiëren van gegevens uit Word (bv. de voorbeelden die je via WordSmith hebt verkregen en die je verder wilt analyseren in Access), als hij in het eerste veld verschijnt. Daarom is het nodig die primaire sleutel te verhuizen naar het tweede veld. Hoe je dit moet doen wordt dadelijk uitgelegd. Je kunt eerst de tabel verder aanmaken.

Vooraf eventjes een aantal spelregels opsommen:

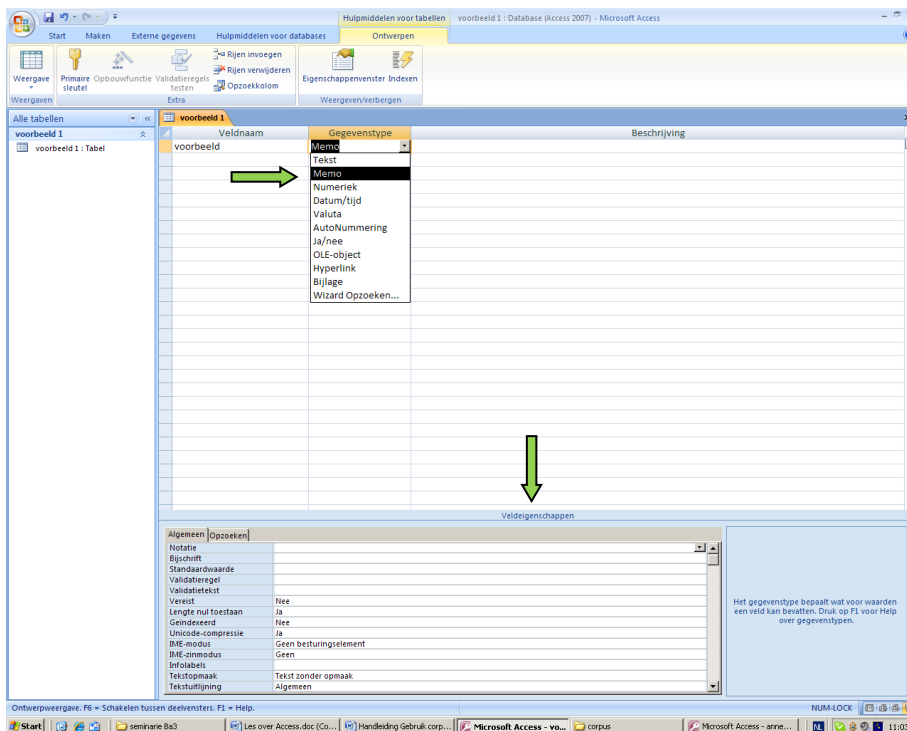
- **Elk criterium**, elk aspect dat je wil onderzoeken in je voorbeelden, stemt overeen met **één rij** in de ontwerpweergave van je tabel. Stel dat je voor *aboutir à* en *déboucher sur* de volgende zaken wil verifiëren: gebruik van *à/sur* (ja of nee), locatieve interpretatie (ja of nee), type onderwerp/X (SN, INF, COMPL), lexeem in de positie van X etc., dan zul je voor elk van deze criteria een rij moeten definiëren.
- In de eerste kolom in de ontwerpweergave zul je de veldnaam moeten definiëren. Dat wil zeggen dat je een naam moet kiezen voor de eigenschap die je wil onderzoeken, bv. "sur/à", "locatief", "type X" of "Xarg", "Xlex"... . Je mag hier naar eigen wens afkortingen gebruiken. Zorg er wel voor dat je achteraf (dagen, weken) nadat je de tabel hebt ontworpen en ze opnieuw opent, je je nog herinnert welk criterium die afkorting precies weergeeft.
- Er bestaan verschillende types criteria (numerieke gegevens wanneer je iets wilt tellen, bv. het aantal woorden in een zin/leeftijd of geboortjaar van een deelnemer aan een

enquête ; etiketten zoals SN, INF, COMPL, ... die eigenlijk van tekstuele aard zijn, etc.). Je moet het juiste type selecteren of invoeren in de tweede kolom.

- In de derde kolom, kun je wat uitleg toevoegen: je kunt bv. de volledige naam invoeren van het criterium dat je onderzoekt als je in de eerste kolom met afkortingen werkt, je kunt er ook de etiketten opsommen die gebruikt zullen worden om het resultaat van de analyse van dat specifiek criterium weer te geven, bv. “SN”, “INF”, “COMPL”, ...

## Stap 1

De allereerste stap die je moet ondernemen, bestaat erin een plaats te voorzien in je tabel voor de voorbeelden die je via WordSmith hebt verkregen. Normaliter heb je die voorbeelden verzameld in een Word document. Voordat je die voorbeelden kunt kopiëren naar Access, maak je in de ontwerpweergave een ‘criterium’ of beter gezegd een ‘veld’ (we nemen vanaf nu de terminologie van Access aan) *voorbeeld*. In het eerste vakje van de tweede rij (eerste kolom), typ je ‘voorbeeld’ of “exemple”. Plaats daarna je cursor in de tweede kolom en klik met de linkermuisknop. Er verschijnt automatisch “tekst” en een pijltje van een rolmenu. Klik op het pijltje. De verschillende types gegevens waartussen je kunt kiezen, verschijnen. Aan elk type zijn een aantal eigenschappen verbonden. Zo is “tekst” beperkt in lengte, het maximaal aantal karakters dat kan worden ingevoerd, is beperkt tot 255. Aangezien de voorbeelden die je wil analyseren, in de meeste gevallen langer zijn dan 255 karakters, is het nodig om “memo” te selecteren. “Memo” laat toe om tekstuele gegevens in te voeren die langer zijn.



Onderaan zie je het deelvenster “veldeigenschappen”. Hier kun je in principe verder de eigenschappen definiëren van het veld, in dit geval het veld “voorbeeld”. Voor dit veld (= het voorbeeld) hoeft je evenwel niets te wijzigen.

## **Stap 2**

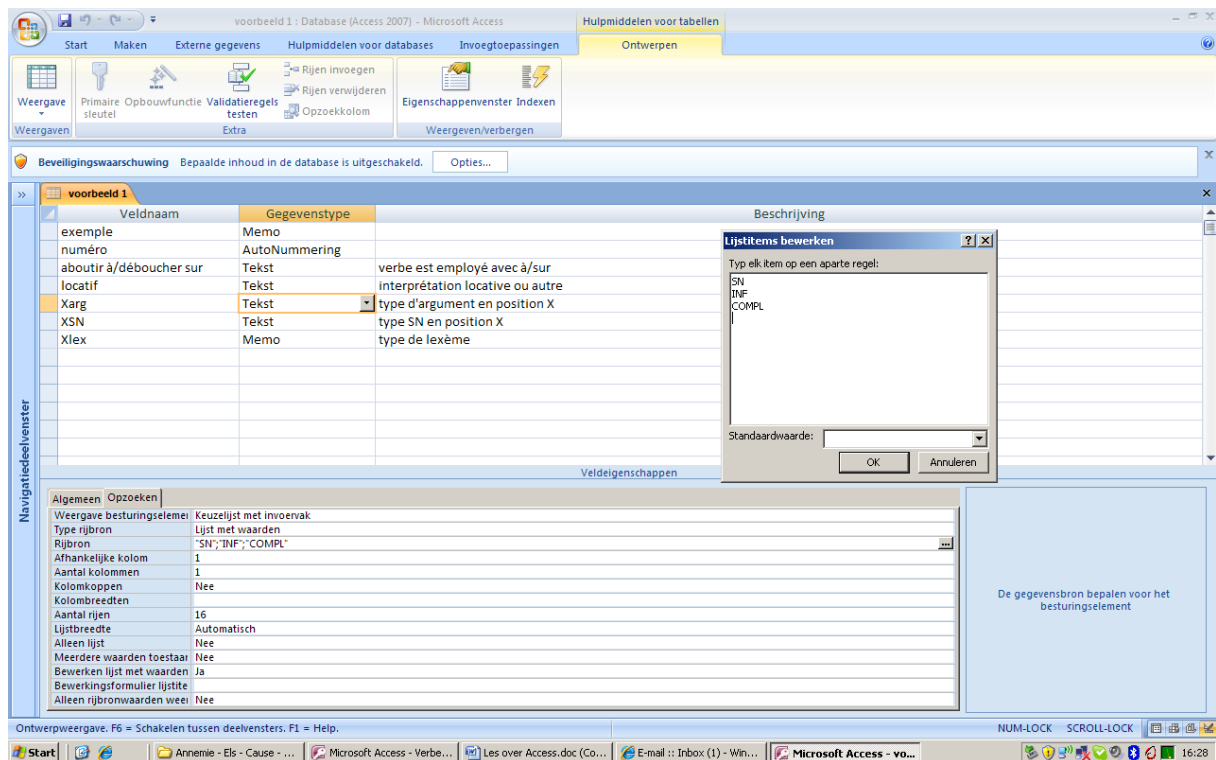
Het is altijd een goed idee om een nummer toe te kennen aan elk voorbeeld. Als je opzoeken doet in je database op een later punt, dan kunnen de rijen (de voorbeelden) in de gegevensweergave van de tabel, een andere volgorde aannemen. Als je tijdens het analyseren iets opmerkt bij één bepaald voorbeeld, dan kun je het nummer dat je eraan hebt toegekend noteren en bijhouden. Dat kan handig zijn als je een aantal moeilijke, speciale of interessante voorbeelden wilt bespreken met je promotor. Dankzij het nummer kun je ze snel terugvinden, want het toegekende nummer blijft hetzelfde, ook al verandert het voorbeeld van plaats in de tabel! Hiervoor kan het veld “Id” worden gebruikt (cf. supra), dat Access automatisch heeft aangemaakt (eerste vakje, eerste rij, eerste kolom).

## **Stap 3**

Vul je tabelontwerp verder aan met de criteria die je wil onderzoeken. Onthoud: voor elk criterium dat je wil gebruiken, vul je de naam in de eerste kolom in; daarna definieer je om welk type gegeven het gaat in de tweede kolom:

- tekst: kies je wanneer je tekst wil invoeren (bv. SN, INF, COMPL) voor elk voorbeeld dat je analyseert (opgelet : het aantal karakters is beperkt tot 255). Dit is een gegevenstype dat je frequent zult gebruiken. Je hebt het nodig bij het eigenlijke analyseren, voorbeeld per voorbeeld (bv. als je wil weten met welke soorten onderwerpen (inf, compl, SN) *aboutir* en *déboucher* voorkomen, als je wil weten welke woorden exact gebruikt worden om de collocaties te bestuderen, als je wil nagaan of, in het geval het onderwerp een SN is, het gaat om een voornaamwoord/eigenaam/ determinant + N etc.)

In sommige gevallen is het aantal keuzemogelijkheden (eerder) beperkt. Zo kan het onderwerp van een werkwoord slechts een SN, een infinitief(zin) of een completiefzin zijn. Dit is ook het geval voor de vraag of *aboutir* gebruikt wordt met het voorzetsel *à* en *déboucher* met *sur*: ja of nee. In dat geval kun je zelf een rolmenuutje maken, dat je kunt gebruiken tijdens het analyseren van de voorbeelden in het formulier. In het deelvenster “veldeigenschappen” onderaan, klik je op “opzoeken”. Er verschijnt automatisch: “tekstvak”. Klik op het pijltje ernaast en selecteer “keuzelijst met invoervak”. In de rij daaronder selecteer je daarna “lijst met waarden”. Zet je cursor vervolgens op de rij daaronder (rijbron) en klik op de drie puntjes ... die rechts verschijnen. Er opent zich een venstertje: je typt er de verschillende opties waaruit je kunt kiezen tijdens je analyse [bv: (ja, nee) of (SN, INF, COMPL)]. Voor elk item ga je naar een nieuwe regel > enter nadat je een item hebt ingevoerd. Laat het vakje naast standaardwaarde leeg. Als de enige opties voor een analyse diegene zijn die je hebt ingevoerd, dan kun je nog de “nee” veranderen in “ja” naast “alleen lijst” iets verder eronder.



- memo: moet je kiezen als de tekst die je wil invoeren langer dan 255 karakters is
- numeriek: moet je kiezen als je een getal wilt invoeren (bv: het aantal woorden tussen een anaforsch gebruikt voornaamwoord en het antecedent, de leeftijd van deelnemers aan een onderzoek, het aantal talen dat iemand spreekt, ...)
- autonummering: Access kent automatisch een nummer toe, te beginnen bij 1 t.e.m. het aantal voorbeelden dat er zijn.

De andere gegevenstypes heb je in principe niet nodig.

In de derde kolom kun je extra informatie voor jezelf toevoegen. Stap 3 zal een beetje zoeken en uitproberen vergen. Het is interessant om zoveel mogelijk criteria bij aanvang aan je tabel toe te voegen, maar je kunt op elk ogenblik tijdens het analyseren of erna terugkeren naar de ontwerpweergave van je tabel en wijzigingen aanbrengen of criteria toevoegen. Natuurlijk betekent elke wijziging dat je elk voorbeeld van voor af aan opnieuw moet bekijken en de analyses moet aanpassen of vervolledigen.

Sla het uiteindelijke resultaat op (schijfje bovenaan links van het scherm), als je denkt alle criteria te hebben toegevoegd. En schakel nu weer over naar de gegevensbladweergave. Merk op dat de kolommen nu een naam gekregen hebben. Elk veld (of criterium = een rij) in de ontwerpweergave stemt overeen met een kolom in gegevensweergave. De eerste kolom stemt overeen met het veld "Id". Om de voorbeelden te kunnen kopiëren vanuit Word naar de tabel die je net hebt gemaakt, moet het veld "voorbeeld" wel overeenstemmen met de eerste kolom. Je moet dus de huidige eerste kolom (Id) van plaats wisselen met de kolom "voorbeeld". Dit doe je door op de kolom voorbeeld te klikken met de rechtermuisknop, die ingedrukt te

houden en daarna de kolom “voorbeeld” te verslepen met de muis naar de eerste plaats in de tabel. Na het kopiëren vanuit Word, zal elke rij of lijn in de tabel, een ‘record’ in Access-terminen, overeenstemmen met één voorbeeld en alle informatie die je analyseert voor dat voorbeeld kun je terugvinden in dat record, in de verschillende kolommen (één kolom per onderzocht criterium). Hoe je de voorbeelden moet kopiëren uit Word, lees je in de volgende paragraaf.

### 2.3.3 Gegevens importeren

Open het Word document waarin je voorbeelden zijn opgeslagen. Je moet eerst een aantal bewerkingen uitvoeren voor je de gegevens kunt importeren, waardoor een aantal gegevens zullen verdwijnen. Wil je het oorspronkelijk Word document behouden uit veiligheidsoverwegingen (wat aan te raden is), sla dan eerst het geopende document op onder een andere naam. Vervolgens wis je alle kolommen, behalve die waarin de voorbeelden staan (zet je cursor op de kolom, klik op de rechtermuisknop en kies verwijderen).<sup>19</sup> Daarna moet je nog nagaan of elk voorbeeld onmiddellijk met tekst begint. Er mogen geen spaties, diakritische tekens of cijfers in het begin van het voorbeeld staan. Indien dat wel het geval is, verwijder je die het best tot aan de eerste echte letter van het eerste woord van het voorbeeld. Selecteer daarna met de muis de enige kolom die nog overblijft en kopieer ze (control + C // of icoontje onder de schaar bovenaan links).

Ga daarna naar de gegevensbladweergave van je tabel in Access. Zet je cursor in het eerste vakje van de eerste rij. Klik op het pijltje onder het icoon “plakken”, links boven. Kies “toevoegen via plakken”. Bevestig. Het aantal voorbeelden dat je toevoegt, zou moeten overeenstemmen met het aantal voorbeelden in je Word document. Ga dit na: in je oorspronkelijke document zijn de voorbeelden genummerd, je kunt ook opnieuw een kolom toevoegen aan de tabel waarin je enkel de voorbeelden hebt behouden en de voorbeelden nummeren door die kolom te selecteren en dan te klikken op (in het menu start > alinea ) het icoontje dat een opsomming met cijfers weergeeft (het tweede in de bovenste rij van het menu alinea). In de tabel in Access verschijnt naast de voorbeelden een automatisch toegekend nummer in de kolom Id (= de kolom “numéro” in de volgende figuur).

Het is wel duidelijk dat het niet handig is om in de tabel zelf verder te gaan met de analyses: het is te moeilijk om de voorbeelden te consulteren. Het lezen en analyseren van de voorbeelden gebeurt in een formulier. Dat formulier zul je eerst moeten ontwerpen.

#### Opmerking

In sommige gevallen kan het interessant zijn om bepaalde velden van je tabel automatisch te laten aanvullen. Voor het voorbeeld van *aboutir* en *déboucher* kan het nuttig zijn om een veld

---

<sup>19</sup> Je voorbeelden staan dus in een tabel die bestaat uit één enkele kolom, één rij per voorbeeld. Als je analyse een bepaald woord beoogt, dan kun je dit woord omzetten in hoofdletters. Zo springt het beter in het oog in Access. Hoofdletters zijn de enige vorm van markering die zichtbaar blijven in Access (geen bold, italic, kleur).

aan te maken met de naam van het werkwoord (*aboutir* of *déboucher*). Om te vermijden dat je voor ieder voorbeeld zelf “aboutir” of “déboucher” moet typen, kun je de velden automatisch laten bijwerken (aanvullen) met een **bijwerkquery (zie 5.3.)**. In het concrete voorbeeld dat we hier behandelen, zul je er dan wel moeten op toezien dat je eerst de gegevens voor *aboutir* importeert en er de bijwerkquery op toepast en dat je pas daarna de gegevens voor *déboucher* importeert. Nadenken is dus (alweer) de boodschap !

### 2.3.4 Formulier ontwerpen

Je kunt pas een formulier aanmaken als je een tabel hebt ontworpen. Elk formulier is aan een tabel gelinkt. De tabel is eigenlijk de echte gegevensbank. Het formulier is slechts een hulpmiddel om de gegevens beter te visualiseren en makkelijker aan te vullen. Aangezien het formulier gelinkt is aan de tabel, kun je het formulier pas aanmaken als je de tabel hebt opgeslagen en gesloten (klik op het kruisje in het deelvenster)<sup>20</sup> of als de tabel openstaat in gegevensbladweergave.

Er zijn verschillende manieren om een tabel aan te maken: je kunt je eigen formulier ontwerpen zoals je zelf wilt, of je kunt werken met de wizard. Het gebruik van de wizard wordt heel kort uitgelegd. Wij geven de voorkeur aan de andere methode.

#### 2.3.4.1 Een formulier maken met de Wizard

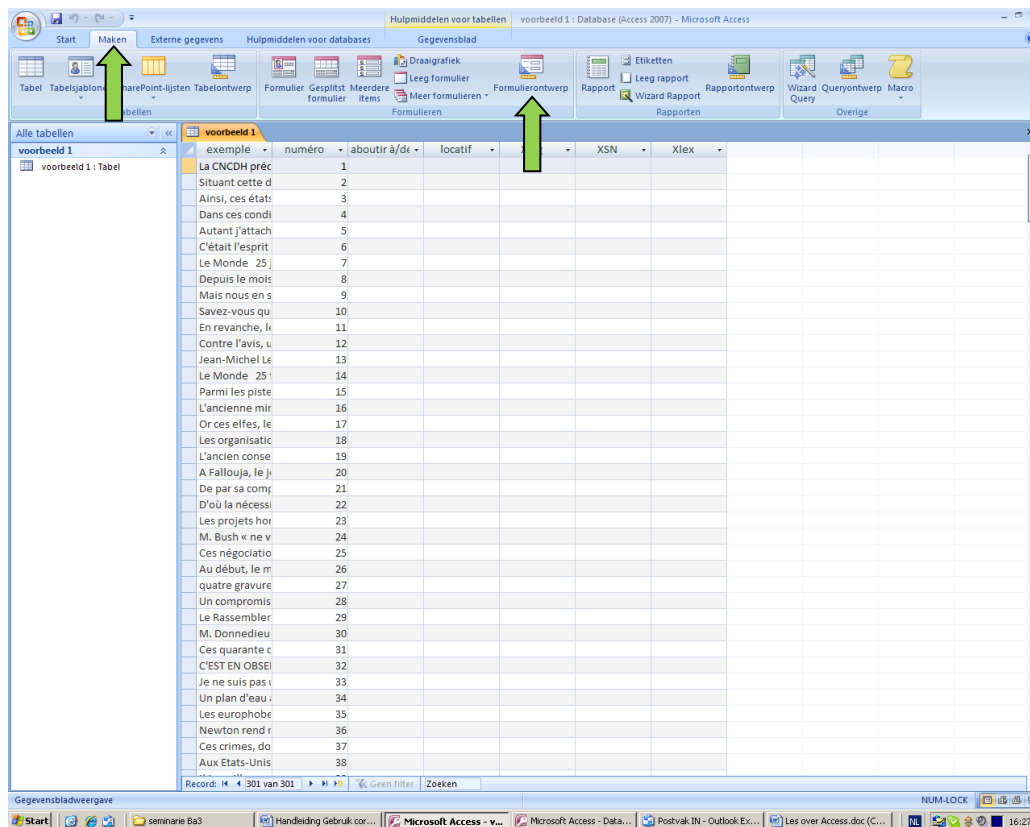
Met de wizard kun je kiezen uit een aantal modellen en ben je gebonden aan het model dat je gekozen hebt. Je vrijheid om te schuiven met je items in het formulier (organisatie en opbouw ervan) is beperkt. Om de wizard te gebruiken, moet je in het menu “maken” en in het submenu “formulieren”, klikken op “meer formulieren”. Er opent een venster. Je selecteert alle velden (>>) [of die velden die je wil analyseren] en klikt op ‘volgende’. Daarna kun je een model kiezen: kolomvorm, tabelvorm, gegevensblad of uitgevuld. Selecteer het model dat je wil en klik op ‘volgende’. Kies daarna de lay-out die je wil en klik op ‘volgende’. Geef het formulier een naam en klik op ‘volgende’. Het formulier is nu klaar. Je ziet de ontwerpweergave (sluit eventueel de tabel als je een melding krijgt dat er een probleem is). Schakel over naar gegevensweergave (links bovenaan het scherm, submenu “weergaven”). Je kunt nu in feite onmiddellijk beginnen analyseren.

#### 2.3.4.2 Een formulier maken via formulierontwerp

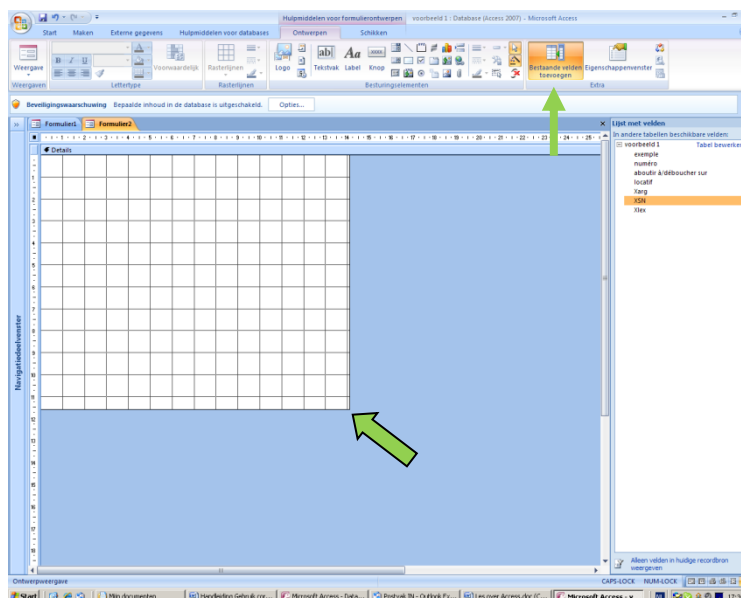
Dit is de verkozen methode. Naast het menu “start” links bovenaan het scherm, zie je het menu “maken”. Klik op “maken”. Klik daarna op het laatste roze icoontje in het submenu formulieren: “formulierontwerp”.

---

<sup>20</sup> Het kruisje bovenaan rechts het scherm dient om het hele programma en de volledige database af te sluiten. Het kruisje bovenaan links in het deelvenster dient om de tabel af te sluiten.



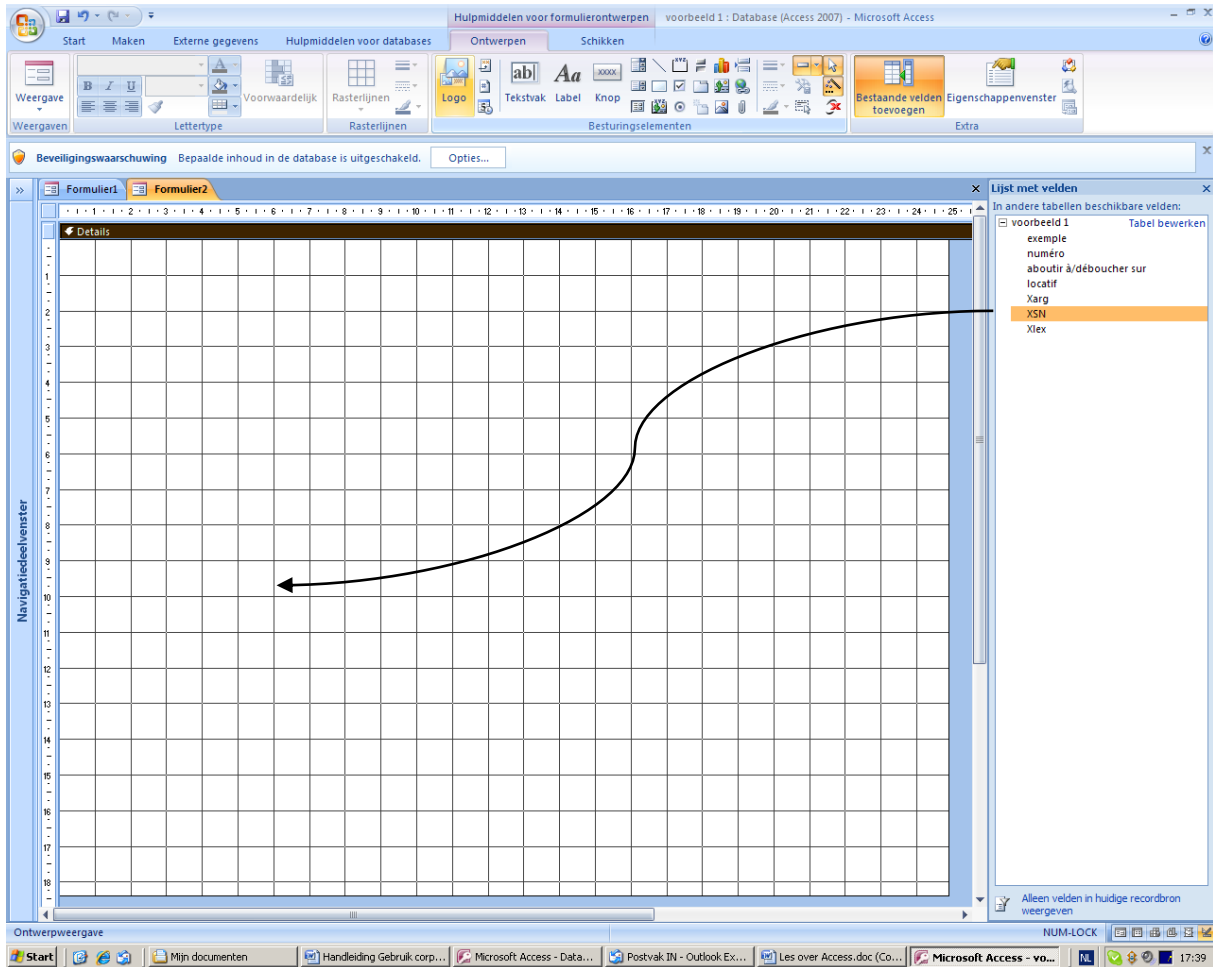
Het formulier opent in de ontwerpweergave. Net zoals bij de tabel, moet je eerst het formulier aanmaken voor je er de gegevens in kunt bekijken en analyseren. Zet je cursor onderaan rechts op de hoek van het raster en versleep de hoek zodat je het raster wat groter kunt maken.



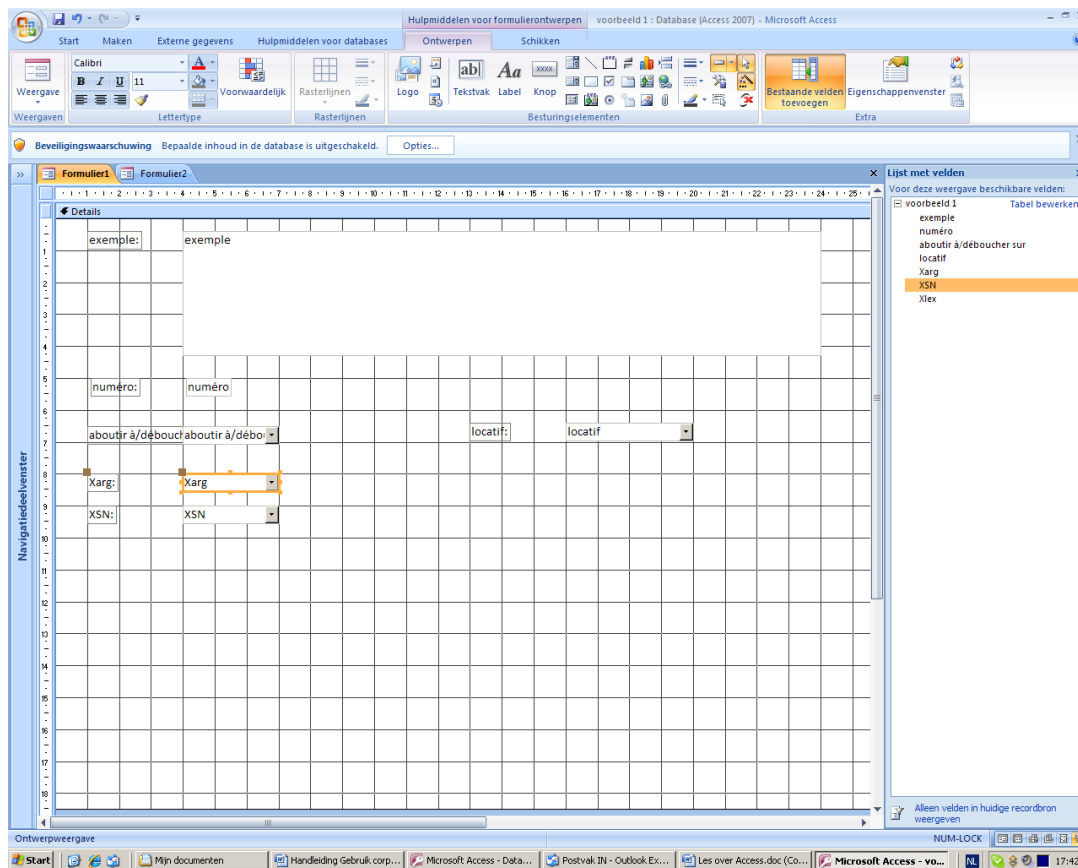
Het menu “ontwerpen” is automatisch geactiveerd. In het laatste submenu “extra” klik je op het eerste icoon “bestaande velden toevoegen” (cf. tweede pijl in de vorige figuur). Je kunt nu elk veld toevoegen aan het formulier en het plaatsen waar je zelf wilt, in de volgorde die je zelf verkiest. Je klikt op het veld dat je wil toevoegen en houdt de linkermuisknop ingedrukt.



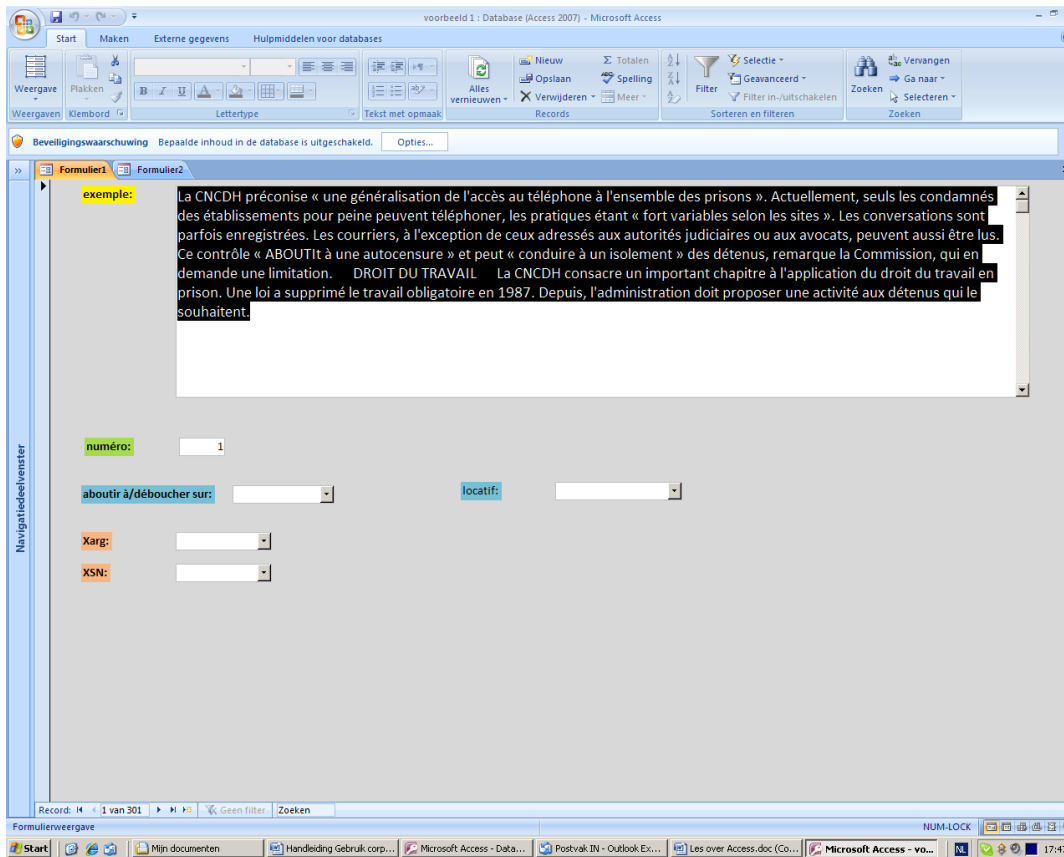
Vervolgens sleep je het veld naar de plaats op je formulier waar je het ongeveer wilt zien verschijnen.



Aangezien je in de ontwerpweergave van de tabel het gegevenstype hebt gedefinieerd, verschijnen er nu tekstvakken, vakken waarin cijfers kunnen worden ingevoerd of rolmenu's, naargelang het type dat je toen hebt gekozen. Je kunt je velden nog steeds verplaatsen en mooi uitlijnen volgens het raster.



Hoe het er uiteindelijk zal uitzien, kun je checken in de gegevensweergave (bovenaan links). Je kunt heen en weer switchen tussen beide weergaves totdat je het gewenste resultaat hebt bekomen. Door op de kadertjes te klikken (opgelet: aanpassingen kun je alleen in de ontwerpweergave aanbrengen!), kun je ze activeren en groter of kleiner maken (zet je cursor op de aangegeven puntjes en sleep totdat je de gewenste grootte bekomt). Je kunt zo ook de afstand tussen het infolabel van het veld en het vakje waarin je iets moet invullen, veranderen. Voorzie een voldoende groot vak voor het voorbeeld, zodat je het in één oogopslag kunt consulteren en niet hoeft te scrollen. Via het submenu "lettertype" kun je de lay-out van je formulier verfraaien. Door je cursor op een willekeurige plaats in het raster te plaatsen en op de rechtermuisknop te drukken, kun je spelen met de achtergrond van je formulier. Ga uiteindelijk naar de gegevensweergave en loop met de tab-toets (de twee pijltjes boven de Shift Lock toets) op je klavier alle velden eens af.



Ga je in de gewenste volgorde van het ene naar het andere veld ? Is het in deze volgorde dat je het voorbeeld wenst te analyseren ? Wijzigingen kun je aanbrengen in de **ontwerpweergave** door op een willekeurige plaats in het raster op je rechtermuisknop te drukken. Eén van de mogelijkheden die verschijnt is “tabvolgorde”. Klik op het veld waarvan je de plaats in de volgorde wilt veranderen en versleep naar de gewenste plaats. Klik op ok. Voor je verder gaat, is het aan te raden om het ontwerp van je formulier op te slaan. Klik op het schijfje bovenaan links en geef het formulier een gepaste naam.<sup>21</sup> Bevestig.

Zorg nu dat je in de gegevensweergave bent. Onderaan kun je zien in welk record (= welk voorbeeld) je aan het werken bent of welk record je aan het bekijken bent. Je begint natuurlijk met het eerste record. Het cijfer onderaan (bv. record 43 van 301), geeft aan dat je eigenlijk in het 43<sup>ste</sup> record (of de 43<sup>ste</sup> rij) van je tabel aan het werken bent (bij elk formulier dat je invult, wordt er een record aan je tabel toegevoegd/verder aangevuld).<sup>22</sup> Om naar het vorige record terug te keren kun je op het enkele pijltje ◀ klikken, om naar het allereerste record terug te keren klik je op ◀◀, om naar het volgende te gaan klik je op ▶ en om naar het laatste te gaan klik je op ▶▶. Wil je terugkeren naar een specifiek record, dan kun je o.m. het cijfer

<sup>21</sup> In oudere versies van Access is het aangeraden om F voor de naam te plaatsen om aan te geven dat het om een formulier gaat.

<sup>22</sup> Bij het uitvoeren van queries of het opzoeken in de formulieren, kan zoals eerder al werd aangegeven, de volgorde van de records veranderen. Noteer daarom het nummer dat in het formulier verschijnt in het veld “nummer” of “numéro” (naar gelang de naam die je in het ontwerp van je tabel aan dit veld hebt gegeven) als je iets over dit bepaalde voorbeeld wil bespreken of noteren.

intikken, in plaats van 43 tik je bv. 26 of je gebruikt de functie “ga naar” in het submenu “zoeken” in het “start” menu. Nog andere mogelijkheden om bepaalde voorbeelden terug te vinden, worden in de volgende sectie uitgelegd (cf. formulierfilter).

Zo, nu ben je in staat om voorbeeld per voorbeeld te analyseren. Vul de cijfers of woorden/afkortingen in de aangewezen velden (vakjes) in of kies het juiste item uit de rolmenu’s (opgelet: analyseren kan enkel in de gegevensweergave).

Je hoeft in Access eigenlijk niet echt tussentijdse resultaten op te slaan zoals bv. in Word. Zodra je naar een volgend record overgaat (via de tab-toets of door onderaan het formulier op het juiste pijltje te klikken), wordt alles wat je voor het vorige hebt ingevuld automatisch opgeslagen. Als je klaar bent of even wilt stoppen, kun je het formulier gewoon sluiten door op het kruisje te klikken (x) bovenaan rechts in het deelvenster.<sup>23</sup> Als je het formulierontwerp nog niet hebt opgeslagen, vraagt het programma je om dat nu te doen.

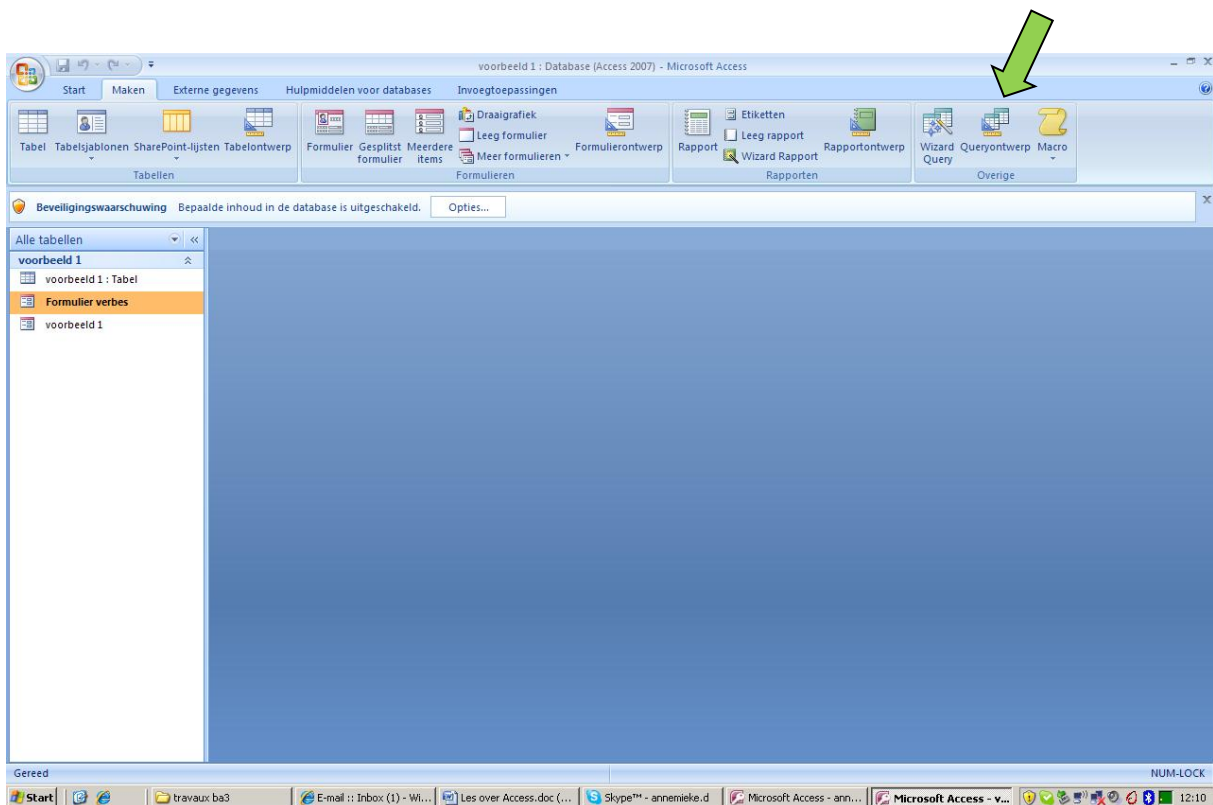
### 2.3.5 Queries ontwerpen

Eens je alle voorbeelden hebt onderzocht, kun je proberen om na te gaan of je tendensen en patronen in je data kunt terugvinden. In veel taalkundig onderzoek betekent dat dat je de frequentie van een aantal dingen nagaat: bv. hoe vaak is X in het geval van *aboutir* een infinitief ? Dit soort simpele berekeningen kun je ook met Excel uitvoeren natuurlijk. Access laat, in tegenstelling tot Excel, toe om veel complexere berekeningen uit te voeren, waarbij verschillende criteria tegelijkertijd betrokken zijn: bv. voor alle Y die SN zijn, die déN zijn en waarvan het lexeem “résultat” is, wil ik weten hoeveel keer het werkwoord in de indicatif en in de PC staat etc. Welke complexe opzoeken zinvol zijn, moet je zelf beslissen. Welke opzoeken mogelijk zijn, hangt natuurlijk af van wat je in je tabel gestopt hebt, met andere woorden van het aantal velden dat je hebt gedefinieerd. Een database aanmaken betekent dus dat je in het begin een beetje speculeert over wat eventueel van belang is. Soms onderzoek je criteria die achteraf niet zo bruikbaar blijken, maar dat is vaak beter dan tijdens het werken met queries vast te stellen dat je niet over voldoende informatie beschikt en dus de ontwerptabel moet aanpassen en alle voorbeelden opnieuw moet doornemen. Het is afwegen wat er meer tijd in beslag neemt.

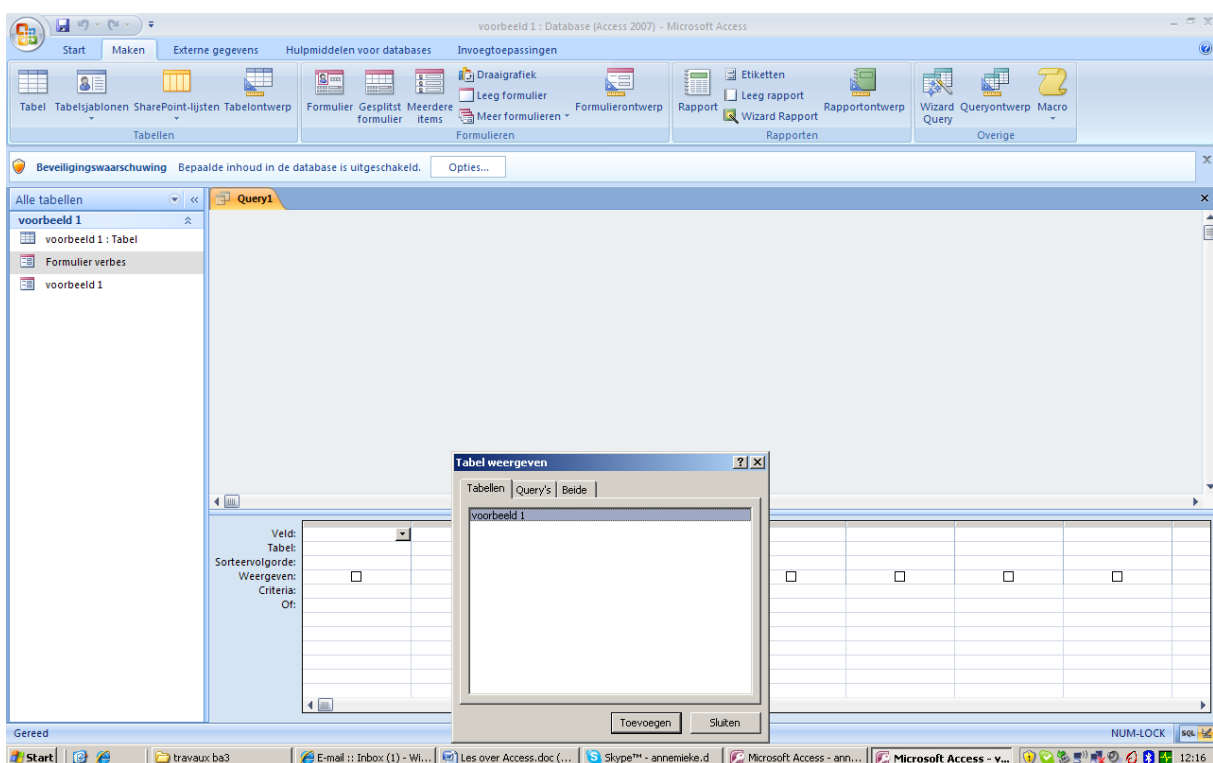
Een query maken doe je door in het menu “maken” te klikken op “queryontwerp”, het tweede icoon in het submenu “overige”:

---

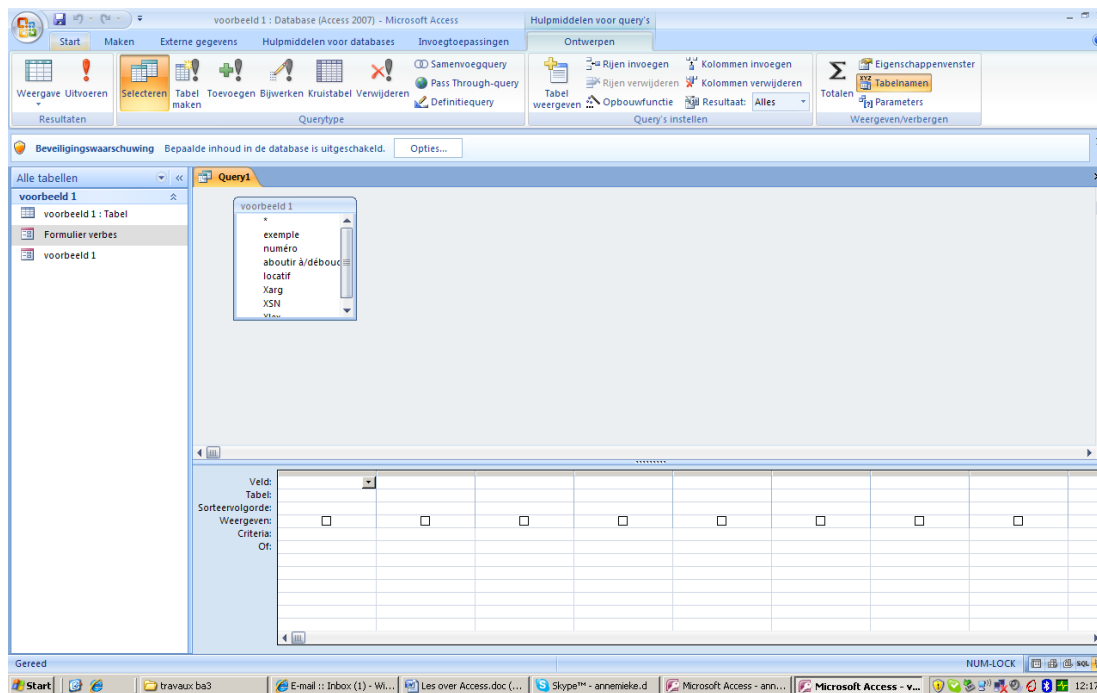
<sup>23</sup> Ter herinnering: het kruisje helemaal bovenaan rechts het scherm dient om het programma af te sluiten.



Er openen twee vensters: één groter deelvenster en een kleiner, apart venster 'tabel weergeven'.



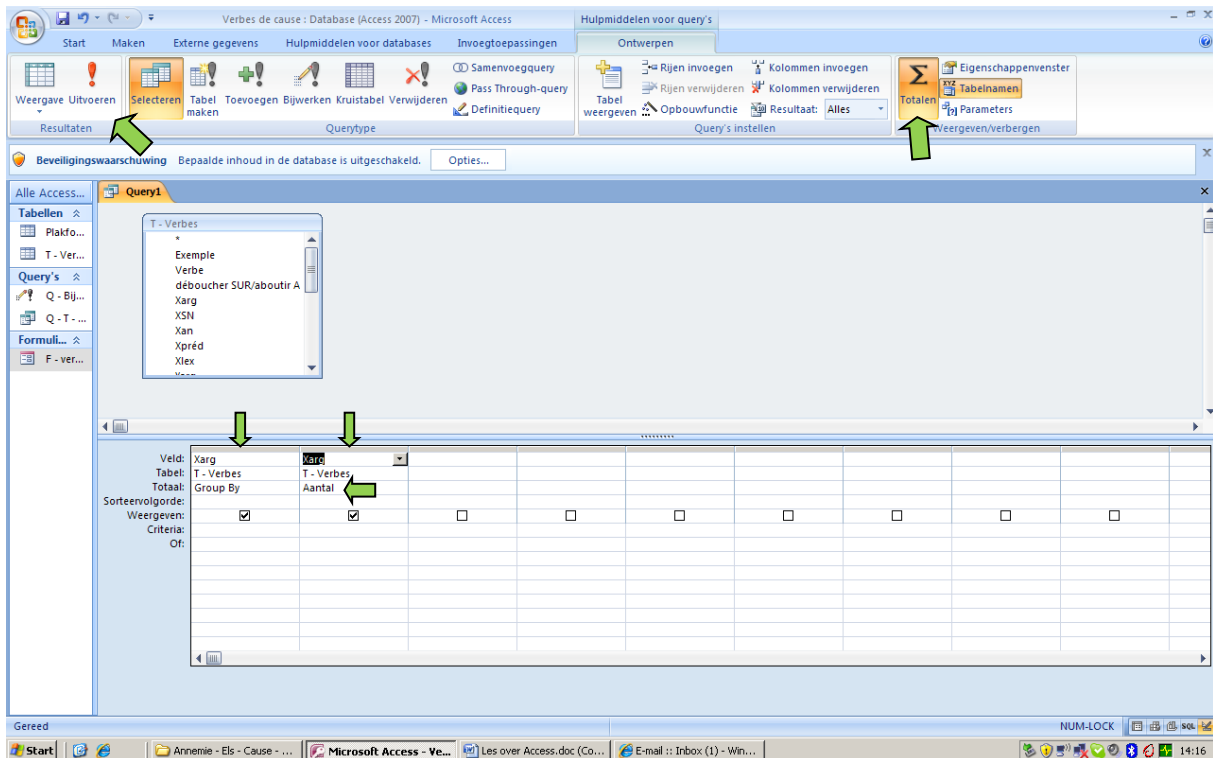
In het kleine venster selecteer je de tabel waarin de gegevens staan waarmee je bewerkingen wilt uitvoeren (normaliter zul je maar één tabel aangemaakt hebben) en je klikt op toevoegen en daarna op sluiten. De verwijzing naar de tabel en de lijst met velden verschijnen nu in het groter deelvenster:



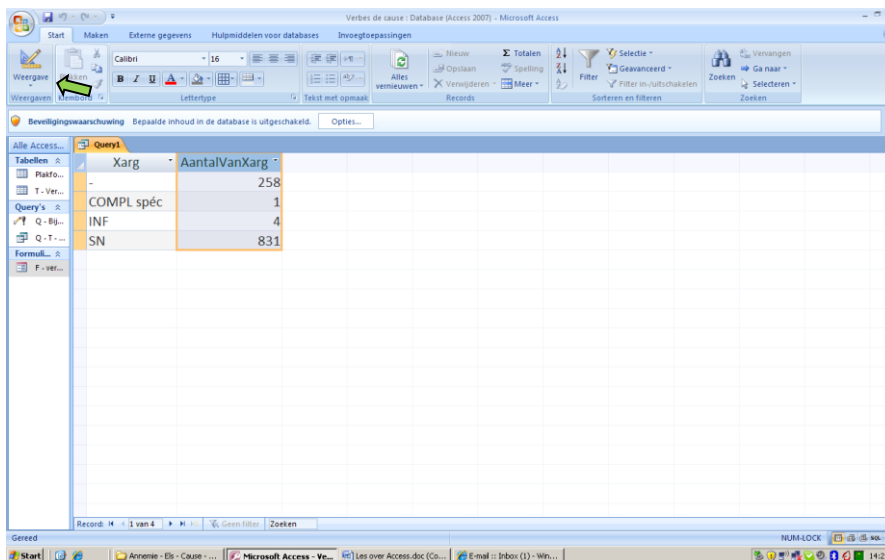
Onderaan het scherm zie je een soort tabel. Het is in deze tabel dat je de zoekopdracht voor Access (de gevraagde bewerking) moet formuleren. In de eerste rij selecteer je via het rolmenuutje het veld waarop je een bewerking wilt uitvoeren. Laten we beginnen met een simpel voorbeeld, daarna proberen we een aantal stappen wat te veralgemenen en uiteindelijk bespreken we een complexer voorbeeld.

### 2.3.5.1 Illustratie

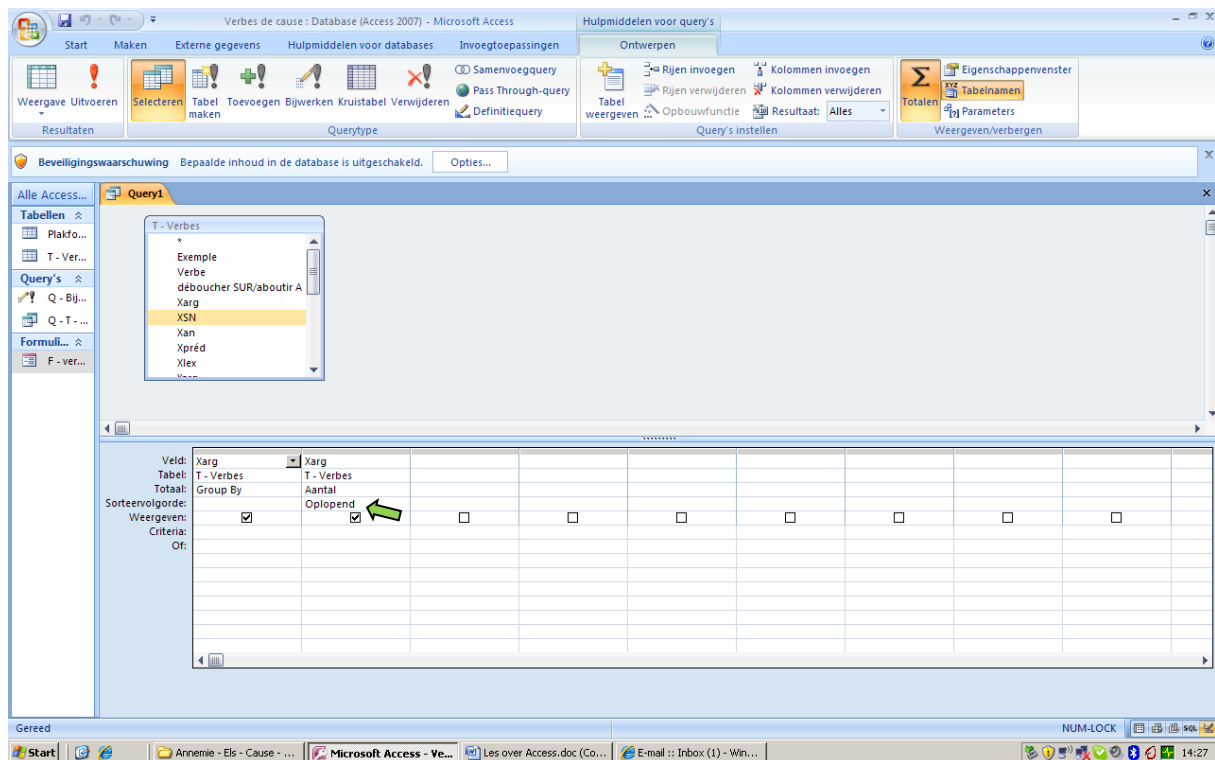
Stel dat je wil nagaan hoe vaak X (het onderwerp van *aboutir/déboucher*) een SN, een INF en een COMPL is. Je kiest als veld (1<sup>ste</sup> rij) : “Xarg” (type argument in X), omdat het bij dat criterium is dat je bent nagegaan of X een SN, INF of COMPL (= completiefzin) is. Doe dit door op het pijltje te klikken en “Xarg” te selecteren. Je hebt nu gedefinieerd welk veld (of welk criterium van je analyse) je verder wil onderzoeken. Nu moet je nog aangeven wat je precies wil onderzoeken met betrekking tot dat veld. In de tweede kolom kies je daarom nogmaals het veld “Xarg”. Daarna klik je op  $\Sigma$  (totalen) in het submenu weergeven/verbergen bovenaan rechts. In de tabel wordt een lijn toegevoegd en er staat automatisch “group by”. In de tweede kolom klik je op “Group by”. Er verschijnt een rolmenu. Vervang “Group by” door op “Aantal” te klikken. Nu heb je verder gedefinieerd dat je het aantal voorbeelden voor elk label (SN, INF, COMPL) in X positie, wil kennen.



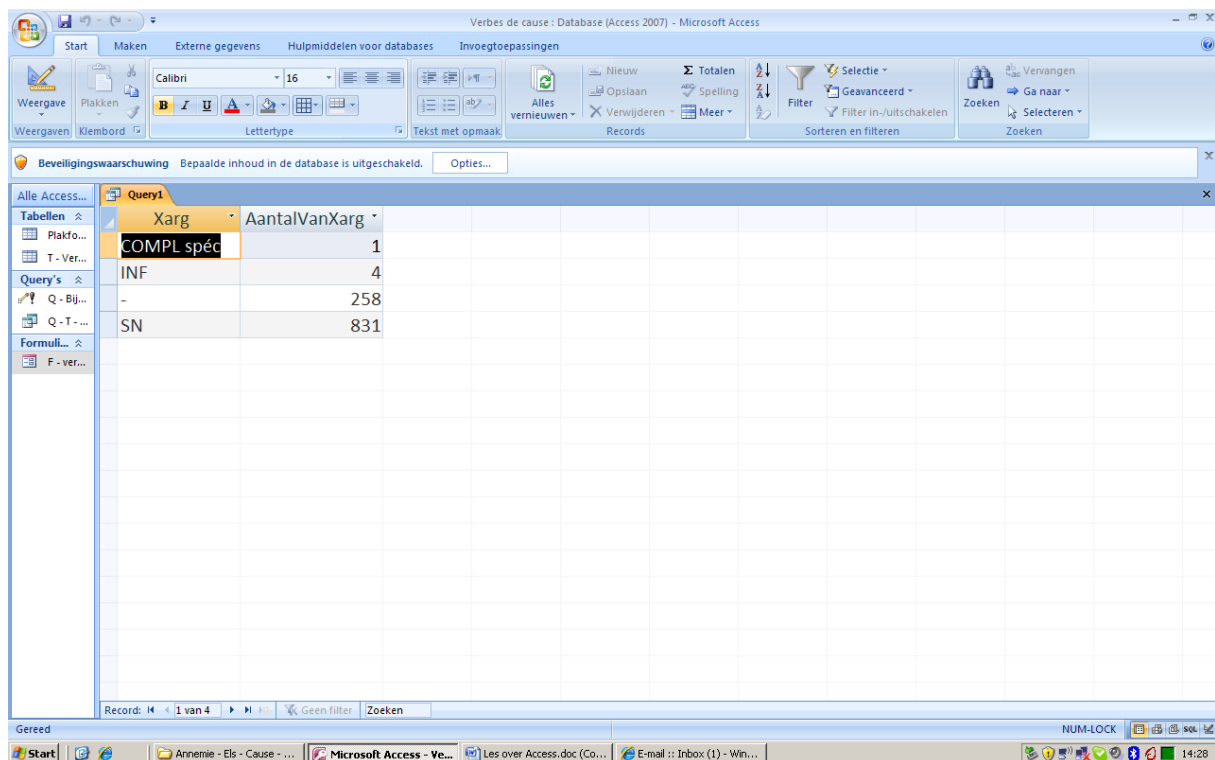
Klik op ! (uitvoeren) in het meest linkse submenu “resultaten”. Nu zie je in de gegevensweergave het resultaat van deze berekening:



Om terug te gaan naar de ontwerpweergave klik je op het weergave-icoontje links bovenaan. Als je de resultaten wat overzichtelijker wil presenteren, dan kun je een sorteervolgorde aangeven. Dit doe je door “oplopend” of “aflopend” te selecteren in de rij “sorteervolgorde” in de ontwerpweergave. Je kunt alfabetisch selecteren op COMPL, INF, SN (ook van Z naar A) of van de hoogste naar de laagste waarde (en omgekeerd) in de tweede kolom.



Klik op uitvoeren ! en bekijk het nieuwe resultaat :

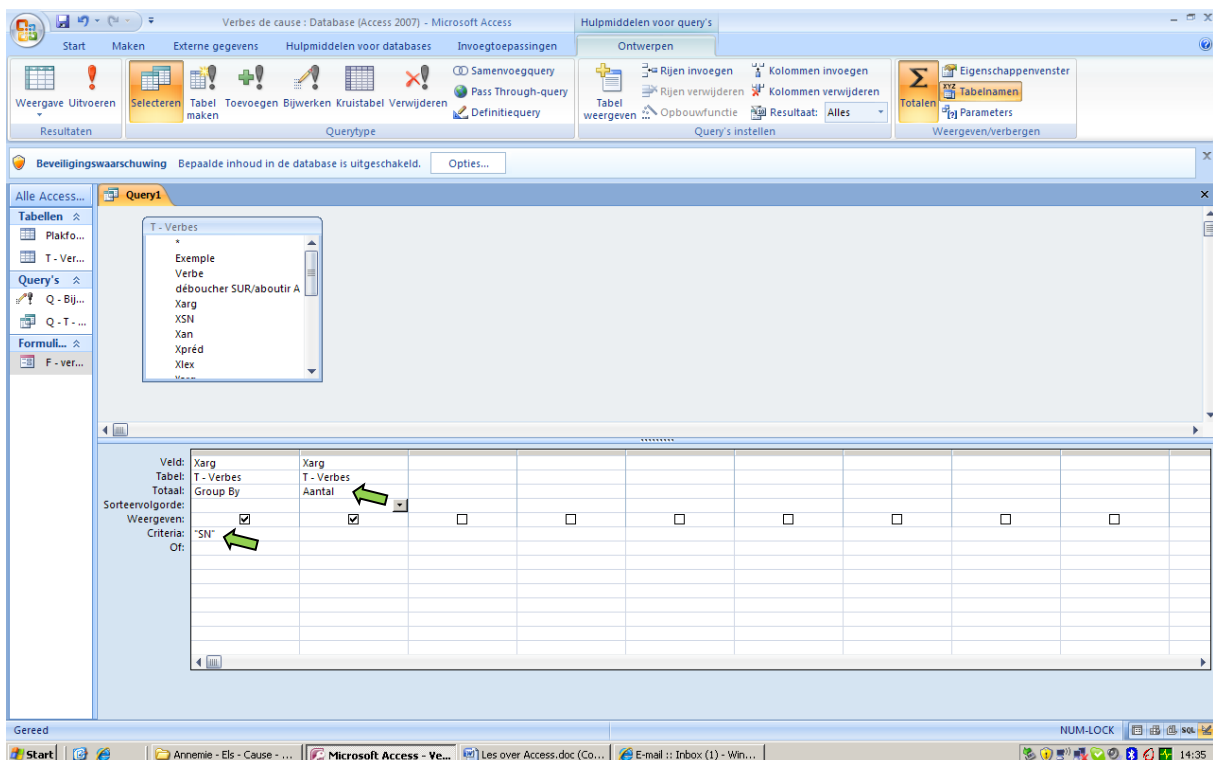




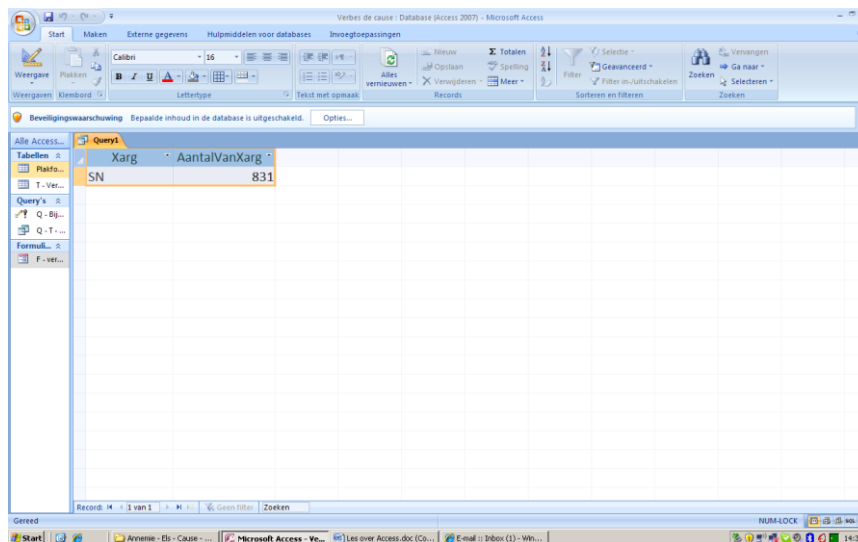
### 2.3.5.2 Mogelijkheden

- Een criterium specificeren

Stel dat je alleen wil nagaan hoe vaak X (het onderwerp van *aboutir/déboucher*) een SN is. Je kiest in het veld voor “Xarg” omdat het bij dat criterium is dat je bent nagegaan of X een SN, INF of COMPL is. In de vijfde rij (criteria) typ je “SN” (er verschijnen automatisch aanhalingstekens, die zijn nodig opdat het programma de opdracht zou uitvoeren, als ze er niet staan, moet je ze toevoegen). Je hebt nu gedefinieerd welk criterium in welk veld je wil onderzoeken. Nu moet je opnieuw aangeven wat je precies wilt onderzoeken voor dat criterium. In de tweede kolom, kies je nogmaals voor het veld Xarg. Daarna klik je op  $\Sigma$  (totalen). In de tabel wordt een lijn toegevoegd en er staat automatisch “group by”. Dit betekent dat alle voorbeelden die voldoen aan de eis dat ze SN zijn in het veld Xarg gegroepeerd worden en de gevraagde bewerking alleen op die voorbeelden zal worden uitgevoerd. In de tweede kolom verander je “Group by” in “Aantal”.



Nu heb je verder gedefinieerd dat je de voorbeelden dat aan dit criterium voldoen, wil tellen. Klik op ! (uitvoeren):



Het is ook mogelijk om meer dan één criterium te specificeren. Misschien wil je enkel gegevens voor COMPL en INF. Dans vul je in de rij criteria in: “COMPL” Or “INF”. Om een selectie van labels/etiketten/waarden te onderwerpen aan een bewerking, gebruik je dus Or om de verschillende elementen tussen aanhalingstekens van elkaar te scheiden. Het is ook mogelijk om na te gaan in welke gevallen/hoe vaak je niets hebt ingevuld in een bepaald veld. Dan vul je in de rij criteria het volgende in: Is null. Omgekeerd kun je ook alle gevallen filteren waarin je wel iets hebt ingevuld, om het even wat, via de expressie: Is not null. Of je kunt specificeren dat je één of meerdere labels/etiketten/waarden wil uitsluiten van de bewerking, bijvoorbeeld : not “SN” of not (“SN” or “INF”). In dat geval gebruik je dus de operator *not* (en niet *Is not*). De waarde of het etiket zet je tussen aanhalingstekens. Als je meerdere etiketten wil uitsluiten, dan som je die op met *or* en zet je het geheel waarop de operator *not* slaat tussen haakjes.

In de rij criteria, geef je (tussen aanhalingstekens) het label of etiket weer dat je gebruikt hebt tijdens je analyses of de waarde (in het geval van cijfergegevens) waarover je iets meer wil weten. Het is dus belangrijk dat je bijhoudt welke labels je hebt gebruikt. In het geval van het rolmenu is dat duidelijk: je hebt de waarden opgegeven bij het ontwerpen van de tabel (bv. SN, INF, COMPL /of/ oui, non). In het geval van een tekstvak, kun je eerst nagaan welke waarden je allemaal hebt ingevuld in de formulieren (bv. in het geval van de lexemen in X en Y positie voor de werkwoorden *aboutir* en *déboucher*). Daarvoor kies je het gewenste veld (bv. Xlex) in de eerste kolom. Je klikt op Σ. Je geeft een sorteervolgorde aan indien gewenst. Door op uitvoeren te klikken, bekom je de lijst met lexemen die in Y positie voorkomen. Nu kun je, indien nodig, verder gaan en één van de lexemen ingeven in de eerste kolom, in de rij “criteria”, om enkel en alleen over dat lexeem meer te weten te komen. Je kunt ook meer dan één criterium invoeren. Als je bv. meer wilt weten over “résultat” en “solution”, dans kun je in de rij “criteria” het volgende invoeren: “résultat” Or “solution”.

- Cijfergegevens bewerken

Als je met cijfergegevens gewerkt hebt in een bepaald veld, dan kun je daar de som en het gemiddelde van berekenen : klik eerst op  $\Sigma$  en selecteer dan de gewenste opdracht in het menu (i.p.v. Group by) : Som of Gem. Je kunt ook vragen wat de laagste ingevulde waarde is (min) of de hoogste (max). Deze bewerkingen kunnen nuttig zijn, als je bijvoorbeeld de gemiddelde leeftijd wil berekenen van de deelnemers aan een enquête, als je wil nagaan hoeveel complementen een werkwoord gemiddeld heeft, of als je wil weten wat het hoogste aantal en het laagste aantal adjectieven is dat een bepaald substantief voorafgaat etc.

- Resultaten sorteren

Door in de rij “sorteervolgorde” voor oplopend of aflopend te kiezen, kun je de manier waarop je gegevens voorgesteld worden, aanpassen. Dit kan het aflezen en interpreteren vergemakkelijken en laat je toe om netjes geordende gegevens naar Excel te exporteren.

- Gegevens verbergen

Tussen de rijen ‘sorteervolgorde’ en ‘criteria’, staat er een rij ‘weergeven’. Je kunt voor elke kolom het vinkje uitschakelen in deze rij. Dit betekent dat je de gegevens voor deze kolom dan niet zult zien in de weergave van de resultaten.

#### 2.3.5.3 De bijwerkquery

Zoals we al vermeldden in hoofdstuk 3 (Gegevens importeren) kan het in sommige gevallen interessant zijn om bepaalde velden van je tabel automatisch te laten aanvullen. Voor het voorbeeld van *aboutir* en *déboucher* kan het nuttig zijn om een veld te hebben met de naam van het werkwoord (*aboutir* of *déboucher*). Dat veld maak je aan als je je tabel ontwerpt (zie hoofdstuk 2). Om te vermijden dat je voor ieder voorbeeld zelf “aboutir” of “déboucher” moet typen, kun je de velden automatisch laten bijwerken (aanvullen) met een **bijwerkquery**. Hoe ga je te werk ?

Voorafgaande opmerking : in het geval dat we hier bespreken, wil je uiteraard voor de voorbeelden van *aboutir* het werkwoord “aboutir” toevoegen en voor de voorbeelden van *déboucher* het werkwoord “déboucher”. Dat betekent hier concreet dat je pas nadat je de bijwerkquery op de voorbeelden voor *aboutir* hebt toegepast, de gegevens voor *déboucher* vanuit Word kunt importeren in Access. Hieronder wordt duidelijk waarom dat zo is.

Om een bijwerkquery te maken, kies je in het menu “maken” op “queryontwerp”. Net zoals voor de “gewone” query’s selecteer je in het venster “tabel weergeven” de tabel waarin de gegevens staan waarmee je de bewerkingen zult uitvoeren en je klikt op “toevoegen” en daarna op sluiten. Opnieuw zie je een scherm met bovenaan de lijst met velden die je aangemaakt hebt en daaronder een soort tabel waar je je opdracht zult formuleren. Nu ga je bepalen dat je geen “gewone” maar wel een bijwerkquery wil maken. Daarvoor klik je in het submenu “Querytype” op “Bijwerken”. Het scherm van daarnet is licht gewijzigd : in de onderste tabel vind je nu een rij “Wijzigen in” boven de rij “Criteria”. Wat we hier willen doen, is bij alle voorbeelden van *aboutir* automatisch in het veld “Verbe” (naam van het

werkwoord) de infinitief “aboutir” laten verschijnen. Je klikt dus in de eerste kolom rechts van “veld” op het pijltje en selecteert daar “Verbe”. Daarna bepaal je dat je voor alle voorbeelden waar het veld “verbe” leeg is, “aboutir” wil zien verschijnen. Je kiest dus als criterium “Is null” en naast “wijzigen in”, typ je “aboutir”. Om de bijwerkquery uit te voeren, klik je op het uitroepteken “uitvoeren” in het submenu “Resultaten”. Access vult nu alle lege “Verbe”-velden aan met “aboutir” en vermeldt in hoeveel gevallen hij dat gedaan heeft : “u staat op het punt x rijen bij te werken”. In principe komt het aantal overeen met het aantal voorbeelden dat je geïmporteerd hebt. Je klikt op “ja” en dan is het bijwerken afgerond.

Je kunt best deze bijwerkquery opslaan.

Laten we nu even het voorbeeld afwerken en “déboucher” erbij nemen. Eens de bijwerkquery toegepast op *aboutir*, kun je de voorbeelden van *déboucher* ook importeren in Access. Als je dat gedaan hebt, open je de opgeslagen bijwerkquery in de linkerkolom van je Acces-venster : je kunt meteen rechts klikken op de opgeslagen query en dan “Ontwerpweergave” selecteren. Je komt weer uit op het scherm met de lijst met velden en de tabel waar je de opdracht invoert. Je ziet dat de vorige opdracht (aboutir) nog ingevuld is. Om nu bij alle voorbeelden van *déboucher* de naam “déboucher” te laten verschijnen in het veld “verbe”, moet je opnieuw aangeven dat je alleen de **lege velden** van “verbe” wil aanvullen (anders overschrijf je “aboutir”). Het criterium blijft dus hetzelfde “Is null”. Naast “wijzigen in” typ je nu “déboucher”. Je klikt op uitvoeren en klaar is kees. Je begrijpt nu wellicht waarom je niet de gegevens voor *déboucher* en *aboutir* samen kon importeren : als je dat had gedaan, dan was het onmogelijk om een onderscheid te maken tussen beide werkwoorden : er zijn immers geen criteria die toelaten te weten in welke voorbeelden het gaat om *aboutir* en in welke andere om *déboucher*.

### Andere mogelijkheden ?

- **Het veld dat je wil bijwerken, hoeft niet leeg te zijn.** Je kunt bepaalde velden ook automatisch wijzigen. Stel dat je in plaats van de Franse benaming “SN” (syntagme nominal) de Engelse benaming “NP” (noun phrase) wil gebruiken. Dan kun je in de bijwerkquery (in ontwerpweergave) het veld “Xarg” aanduiden, naast “criterium” “SN” typen en naast “wijzigen in” “NP” typen. Dit betekent : als “Xarg” gelijk is aan “SN”, vervang dan “SN” door “NP”.
- **Je kunt met meerder kolommen/velden werken.** Stel dat je een veld hebt aangemaakt om naar de bron (*Libération, Le Monde...*) te verwijzen en stel dat je de voorbeelden voor *aboutir* in *Libération* 2005 gezocht hebt en die van *déboucher* in *Le Monde* 2006. De bijwerkquery laat dan toen om, voor elk van beide werkwoorden de bron automatisch aan te vullen. Dat doe je in twee stappen, eerst voor het ene werkwoord en dan voor het tweede : 1) Voor *aboutir* : in de eerste kolom selecteer je het veld “verbe” ; in de tweede kolom het veld “Bron/source”. In de eerste kolom typ je “aboutir” naast “criterium” ; in de tweede kolom typ je “Libération 2005” in de rij van “wijzigen in”. Je klikt op uitvoeren en zo wijzigt Access voor alle voorbeelden waar “Verbe” = “aboutir” de bron in “Libération

2005”. 2) Voor *déboucher* : je voert dezelfde operatie nog eens uit, maar nu met “déboucher” naast “criterium” voor het veld “Verbe” en “Le Monde 2006” naast “wijzigen in” voor het veld “bron/source”.

#### 2.3.5.4 Complexe zoekopdrachten: een voorbeeld

Het volgende voorbeeld illustreert de voordelen van Access ten opzichte van Excel en toont aan dat het uitwerken van een query berust op een denkoefening. Stel dat je een corpus hebt samengesteld van voorbeelden van *aboutir* en *déboucher*. Die voorbeelden heb je aan een analyse onderworpen van verscheidene criteria. Je hebt o.m. de volgende zaken onderzocht en ingevuld in een formulier:

- i. Van welk werkwoord het een voorbeeld is: *aboutir/déboucher* (= de waarden voor het veld verbe)
- ii. of het gaat om de constructie *aboutir à* of *déboucher sur*: Y/N (= de waarden voor het veld déboucher SUR/aboutir A)
- iii. of het gaat om een locatieve, een resultatieve of een andere interpretatie: loc/Y (=yes)/NS (= not sure)/- (= niet van toepassing) (= de waarden voor het veld *Y est le résultat*)
- iv. het type argument in X positie: SN/INF/COMPL/- (= de waarden voor het veld Xarg)
- v. het type argument in Y positie: SN/INF/COMPL (= de waarden voor het veld Yarg)
- vi. het type SN in X positie: détN/pro/Npropre/ - (-betekent “niet van toepassing) (de waarden voor het veld XSN)
- vii. het type SN in Y positie: détN/pro/Npropre/ - (-betekent “niet van toepassing) (de waarden voor het veld YSN)
- viii. of de referent in X positie een levend wezen is (animé): Y/N/- (= de waarden voor het veld Xan)
- ix. of de referent in Y positie een levend wezen is (animé): Y/N (de waarden voor het veld YSN)
- x. het lexeem in X positie: je hebt het hoofd van het SN getypt in het enkelvoud (bv. *négociation*) ofwel de inf (die correspondeert met de inf of het ww in de COMPL) > waarden voor het veld Xlex
- xi. het lexeem in Y positie: je hebt het hoofd van het SN getypt in het enkelvoud (bv. *résultat*) ofwel de inf (die correspondeert met de inf of het ww in de COMPL) > waarden voor het veld Ylex

- xii. je hebt de ‘wijs’ en ‘tijd’ geanalyseerd: ind present, inf présent, part présent, ind PC, ind imparfait, ind PS, passé antérieur, futur simple, conditionnel passé, .... > waarden voor het veld *temps + mode*
- xiii. Je bent nagegaan of het ww voorafgegaan werd door een modaal of aspectueel hulpww ; pouvoir, devoir, ... > waarden voor het veld *aux mod/aspect*
- xiv. Je hebt geverifieerd of de zin waarin het ww staat geaffirmeerd of ontkend wordt: aff/nég > waarden voor het veld *aff/nég*

Een voorbeeld van een complexe zoekopdracht kan zijn: geef mij voor alle voorbeelden van *aboutir*, in de constructie met *à*, die geen locatieve interpretatie hebben, en waarvan X een SN is en non animé, de lexemen in X positie en het aantal keer dat elk lexem voorkomt, als het lexem in Y positie “résultat” is.

Kolom 1: veld = verbe, criteria = “aboutir”

Kolom 2: veld = déboucher SUR/aboutir A, criteria= “Y »

Kolom 3: veld = Y est le résultat, criteria: Not “loc”

Kolom 4 : veld = Xarg, criteria = « SN »

Kolom 5: veld = Xan, criteria = “N”

Kolom 6: veld = Ylex, criteria = “résultat”

Kolom 7: veld = Xlex

Kolom 8: veld = Xlex

Klik op  $\Sigma$ . Verander “Group by” in de laatste kolom in “aantal”. Kies oplopend bij sorteervolgorde in kolom 7 om de lexemen in X positie alfabetisch te ordenen. Dat geeft het volgende resultaat in ontwerpweergave en in gegevensweergave:

Verbes de cause : Database (Access 2007) - Microsoft Access

Hulpmiddelen voor query's

Start Maken Externe gegevens Hulpmiddelen voor databases Invoegtoepassingen

Weergave Uitvoeren Resultaten

Selecteren Tabel maken Toevoegen Bijwerken Kruistabel Verwijderen

Querytype

Samenvoegquery Pass Through-query Definitiequery

Ontwerpen

Tabel weergeven Rijen invoegen Rijen verwijderen Opbouwfunctie

Kolommen invoegen Kolommen verwijderen Resultaat: Alles

Query's instellen

Eigenschappenvenster Tabelnamen Parameters Weergeven/verbergen

Beveiligingswaarschuwing Bepaalde inhoud in de database is uitgeschakeld. Opties...

Query1 Q - T - Verbes - QTEL F - verbes

T - Verbes

- \*
  - Exemple
  - Verbe
  - déboucher SUR/aboutir A
  - Xarg
  - XSN
  - Xan
  - Xpréd
  - Xlex
  - Yest

Navigatiedeelvenster

Veld:	Verbe	déboucher SUR/aboutir	Y est le résultat	Xarg	Xan	Ylex	Xlex	Xlex		
Tabel:	T - Verbes	T - Verbes	T - Verbes	T - Verbes	T - Verbes	T - Verbes	T - Verbes	T - Verbes		
Totaal:	Group By	Group By	Group By	Group By	Group By	Group By	Group By	Group By		
Sorteervolgorde:										
Weergeven:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Criteria:	"aboutir"	"Y"	Not "loc"	"SN"	"N"	"résultat"				
Of:										

Gereed

NUM-LOCK SCROLL-LOCK

Start | Annemie - Els - Cause - ... | Microsoft Access - Ve... | Les over Access.doc (Co... | E-mail :: Inbox (1) - Win... | 15:39

Verbes de cause : Database (Access 2007) - Microsoft Access

Start Maken Externe gegevens Hulpmiddelen voor databases Invoegtoepassingen

Weergave Plakken Lettertype Tekst met opmaak

Alles vernieuwen Nieuw Opslaan Verwijderen Records

Totalen Spelling Meer

Filter Selectie Geavanceerd Filter in-/uitschakelen Sorteren en filteren

Zoeken Vervangen Ga naar Selecteren Zoeken

Beveiligingswaarschuwing Bepaalde inhoud in de database is uitgeschakeld. Opties...

Query1 Q - T - Verbes - QTEL F - verbes

Verbe	débouch	Y est le r	Xarg	Xan	Ylex	Xlex	AantalVanXlex
aboutir	Y	Y	SN	N	résultat	action	1
aboutir	Y	Y	SN	N	résultat	choix	1
aboutir	Y	Y	SN	N	résultat	enquête	1
aboutir	Y	Y	SN	N	résultat	interdire	1
aboutir	Y	Y	SN	N	résultat	mélange	1
aboutir	Y	Y	SN	N	résultat	recherche	1
aboutir	Y	Y	SN	N	résultat	réforme	1
aboutir	Y	Y	SN	N	résultat	somme	1
aboutir	Y	Y	SN	N	résultat	test	1

Record: 14 1 van 9 Geen filter Zoeken

Gereed NUM-LOCK SCROLL-LOCK

Start Annemie - Els - Cause - ... Microsoft Access - Ve... Les over Access.doc (Co... E-mail :: Inbox (1) - Win... 15:40



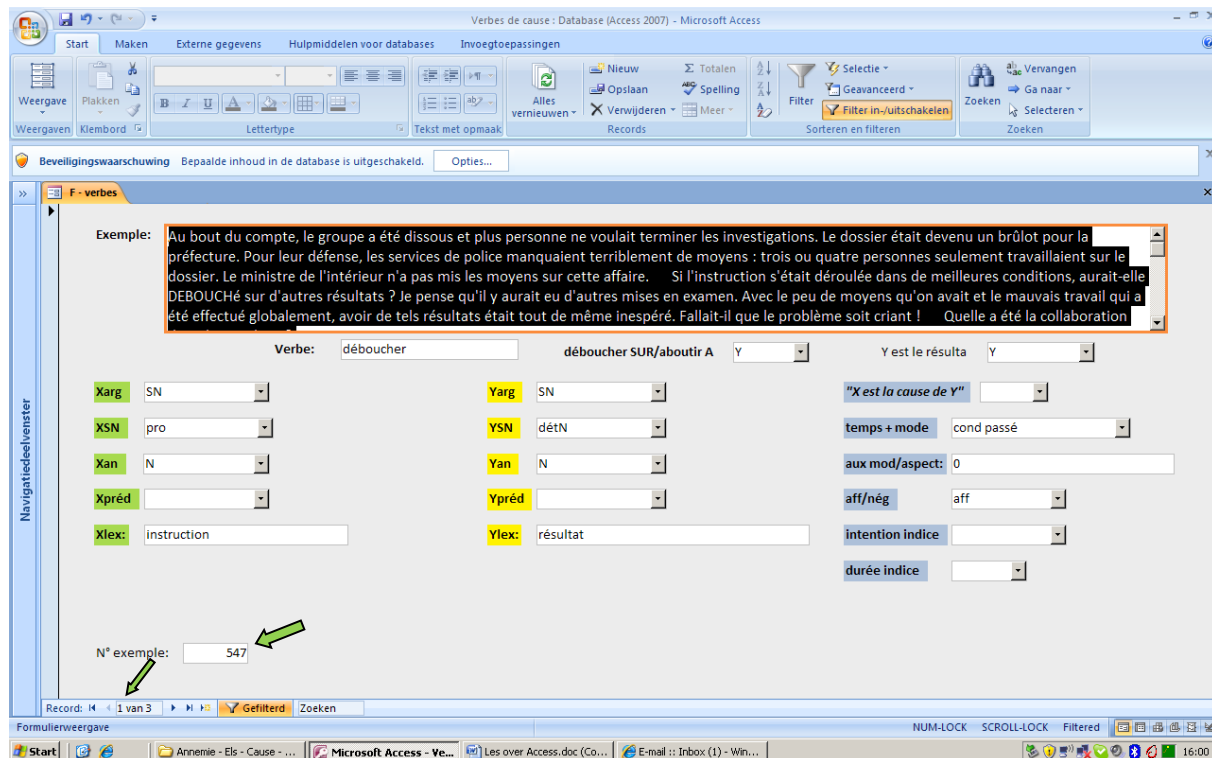
Desgewenst, kun je hier een aantal kolommen verbergen.

### 2.3.5.5 Formulierfilters

Afhankelijk van je doelstellingen, kun je ook opteren voor een zoekopdracht in het formulier zelf. Deze werkwijze heeft als voordeel dat je op basis van een aantal criteria voorbeelden uit het geheel kunt filteren en dat je die voorbeelden nog eens van dichtbij kunt bekijken. Als je werkt met queries, zie je de voorbeelden niet. Anderzijds verkrijgt je, door met een query te werken, een nette tabel als eindresultaat, die je in Excel verder kunt bewerken en in je paper kunt ‘plakken’. Dit kun je dan weer niet via de formulierfilter. Die laat je wel toe een voorbeeld te zoeken dat aan een aantal criteria voldoet. Dat voorbeeld kun je dan op zijn beurt ‘plakken’ in je paper om een bepaald punt te illustreren (de referentie van het voorbeeld zoek je dan achteraf opnieuw op via de cd-rom).

Sluit voor de overzichtelijkheid, alle queries af. Je kunt ze opslaan als je ze later wilt hergebruiken en ze niet opnieuw wilt opstellen. Geef ze een zo duidelijk mogelijke naam. Open opnieuw het formulier. In het submenu “sorteren en filteren” van het “start” menu klik je op “geavanceerd”. Opteer in het rolmenu voor “formulierfilter”. Je krijgt nu een soort leeg formulier te zien. Geen nood: je ingevulde records zijn niet verdwenen! Dit is slechts een soort interface of ‘filter’ waarin je kunt specificeren aan welke criteria de voorbeelden die je wil herbekijken, moeten voldoen. Een voorbeeldje: stel dat je op zoek bent naar een voorbeeld van het lexeem “résultat” in Y positie voor het werkwoord *déboucher*, als illustratie voor je paper. In het veld *verbe* typ je *déboucher*, in het veld *Yarg* typ je *résultat*. Er verschijnen vanzelf aanhalingstekens.

Daarna klik je in het submenu “Sorteren en filteren” op “Filter in-/uitschakelen” en verkrijg je volgend resultaat:



Onderaan zie je “1 van 3”. Dit betekent dat er 3 voorbeelden zijn van *déboucher* waar Y ‘résultat’ is. Je kan via de pijltjes onderaan van het ene naar het andere voorbeeld gaan. De tekst in het veld Exemple kun je selecteren (met de muis) en kopiëren (icoon onder het schartje) en plakken in je Word document.

Dankzij het automatisch nummer dat je hebt toegekend aan elk voorbeeld, kun je via de filter vlug het voorbeeld dat je wou bespreken, oproepen en kopiëren, zodat je het aan je promotor kunt laten zien.

Je kunt dus duidelijk heel verschillende zoekbewerkingen uitvoeren. Nog even opmerken dat je ook gebruik kunt maken van het jokerteken \* om bepaalde woorden af te korten of om duidelijk te maken dat wat volgt, voorafgaat of er tussenin staat (naargelang de plaats van de asterisk(en)) niet van belang is. Verder kun je ook hier werken met de expressies *Is null* of *Is not null*.<sup>24</sup>

De filter opheffen om terug te keren naar het geheel van voorbeelden doe je door opnieuw op “Filter in-/uitschakelen” te klikken. Filters die je hebt gecreëerd, kun je wissen door op het icoon “geavanceerd” te klikken en dan op “alle filters wissen”.

<sup>24</sup> Als je twijfelt over de analyse van bepaalde criteria, gebruik dan geen ? vraagteken, maar gebruik bv. d (doute), ps (pas sur), ... Je kunt namelijk niet zoeken op vraagtekens in Access 2007 (wel in vroegere versies).

### **2.3.6 Resultaten exporteren**

De resultaten in de gegevensweergave van de query, kun je selecteren met de muis en kopiëren (icoon net onder de schaar, naast plakken in het submenu “klembord” van het “start” menu). Open dan een Excel bestand, plaats je cursor waar je de tabel met resultaten wilt hebben en klik op plakken. Nu kun je er verder bewerkingen op uitvoeren in Excel: percentages berekenen, diagrammen of grafieken maken, de lay-out van de tabel aanpassen (kleuren, lettertypes, etc.) ...

## 2.4 Gegevens samenvatten met draaitabellen<sup>25</sup>

Bij taalkundig corpusonderzoek is het typisch dat je uiteenlopende factoren annoteert en analyseert van één of ander taalkundig verschijnsel. Die gegevens kun je overzichtelijk opslaan op in één of andere databank met behulp van software zoals Access of Filemaker (cf. 2.3).

Na de annotatiefase wil je de vele verzamelde gegevens samenvatten in overzichtelijke tabellen en grafieken, om zo trends en tendensen in het materiaal te ontdekken. Een handig hulpmiddel hierbij zijn de zogenaamde **draaitabellen** of, in het Engels, **pivot tables**. Een draaitabel is een dynamisch samenvattingsrapport gemaakt uit een databank of een tabel.

Draaitabellen vind je in rekenprogramma's (bv. Excel, Open Office, Lotus 1-2-3), in statistische programma's (bv. SPSS) en in databanksoftware (bv. Access). Met behulp van draaitabellen kan je data uit een basistabel automatisch sorteren en tellen, en op basis daarvan een tweede tabel creëren waarin de samengevatte data in de vorm van een kruistabel is weergegeven. Het samenvatten van de data is erg intuïtief, waarbij de data uit de basistabel grafisch versleept kan worden in een samenvattende kruistabel. Dat grafisch verslepen en verdraaien van data tot een tabel heeft aanleiding gegeven tot de naam *pivot table* of draaitabel.

Wat is het voordeel van een draaitabel:

- dynamische weergave: aanpassen rijen, kolommen, soort frequentie
- verbonden met databank > bij updaten van de databank, wordt ook de draaitabel aangepast
- snel en intuïtief

In deze bijdrage demonstreren we kort hoe je de draaitabelfunctie uit Excel 2007 kunt gebruiken. Meer informatie over draaitabellen vind je ook op de volgende webadressen:

[http://www.gratis cursus.be/excel\\_2007/excelles047.htm](http://www.gratis cursus.be/excel_2007/excelles047.htm)

<http://www.soweb.be/office2007/excel2007/draaitabellen.pdf>

Voor meer informatie over hoe je draaitabellen maakt in Access en SPSS, verwijzen we naar de volgende webpagina's:

Access: [http://www.gratis cursus.be/access\\_2007/AccessLes061.htm](http://www.gratis cursus.be/access_2007/AccessLes061.htm)

SPSS: [http://www.ats.ucla.edu/stat/Spss/library/sp\\_pivot.htm](http://www.ats.ucla.edu/stat/Spss/library/sp_pivot.htm)

### Stap 1

---

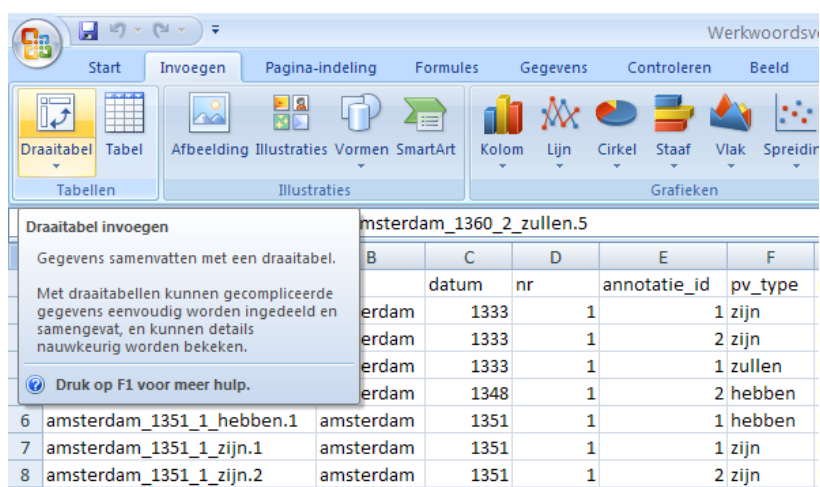
<sup>25</sup> Red.: Evie Coussé, november 2009.

Open je gegevens in Excel. Je kan zowel databestanden importeren uit een andere databank of uit een tekstbestand waarin de kolommen via een scheidingsteken (bv. tab) zijn gemarkeerd.

Zorg ervoor dat elke kolom een unieke kolomtitel krijgt (bv. 'datum', 'pv\_type'); zodat je later in de draaitabel gemakkelijk de gegevens uit die kolom kunt herkennen en oproepen.

## Stap 2

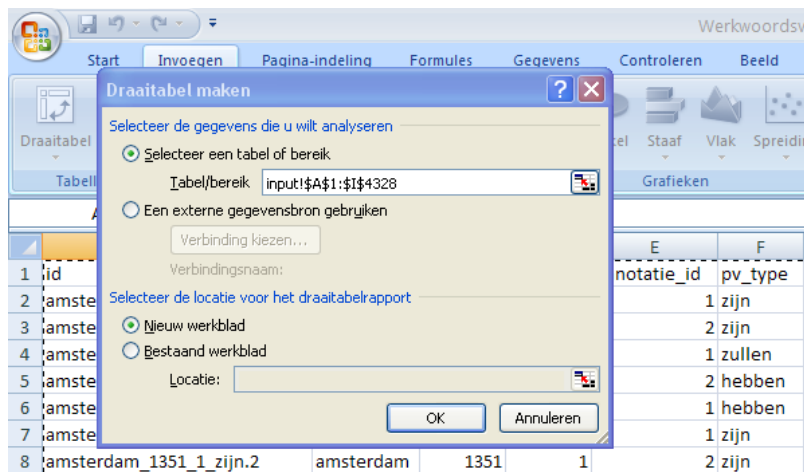
Ga naar het menu 'invoegen' en kies daar de optie 'draaitabel' links van de keuzemogelijkheden.



## Stap 3

Excel zal een pop-up venster openen waarin je de gegevens kunt selecteren die je wilt samenvatten. Normaalgezien zal Excel zelf alle ingevoerde gegevens voor je selecteren. Mocht je om één of andere redenen slechts een selectie van je gegevens willen samenvatten (bv. om geheugen te sparen) dan kun je manueel ook een ander gegevensbereik invoeren.

Zorg ervoor dat je draaitabel op een nieuw werkblad terechtkomt. Dat maakt je Excel-bestand achteraf overzichtelijker. Het is ook handig om je nieuwe werkblad een betekenisvolle naam te geven, zodat je achteraf precies weet wat er op elk werkblad staat.



## Stap 4

Je nieuwe werkblad toont aan de rechterkant een lijst met draaitabelvelden. Door een aantal van die velden te selecteren en naar het onderstaande kolom- of rijgebied te slepen (hou de rechtermuisknop ingedrukt) kan je de draaitabel aan de linkerkant vullen.

In onderstaand voorbeeld is het veld 'periode\_25jaar' naar het rijgebied gesleept, het veld 'wwe' naar het kolomgebied gesleept en is het veld 'id' naar het waardegebied gesleept. Mocht je geen veld naar het waardegebied slepen, dan blijft je draaitabel zonder frequenties. In het voorbeeld wordt de som gemaakt van de elementen per cel in de tabel. Je kan er ook voor kiezen om je gegevens op een andere manier samen te vatten, door bijvoorbeeld gemiddeldes of medianen.

Werkwoordsvolgorde.xlsx - Microsoft Excel

Start Invoegen Pagina-indelii Formules Gegevens Controleren Beeld Opties Ontwerpen

Draaitabel Actief veld Groeperen Sorteren Vernieuwen Andere gegevensbron Acties Extra

Lijst met velden Knoppen +/- Veldkopeteksten Weergeven/verbergen

A3 Aantal van id

	A	B	C	D	E
2					
3	Aantal van id	Kolomlabels			
4	Rijlabels	groen	rood	Eindtotaal	
5	1250-1274	83	50	133	
6	1275-1299	286	172	458	
7	1300-1324	119	47	166	
8	1325-1349	145	29	174	
9	1350-1374	180	29	209	
10	1375-1399	236	31	267	
11	1400-1424	274	16	290	
12	1425-1449	298	27	325	
13	1450-1474	292	40	332	
14	1475-1499	292	21	313	
15	1500-1524	326	28	354	
16	1525-1549	220	35	255	
17	1550-1574	195	52	247	
18	1575-1599	84	32	116	
19	1600-1624	55	35	90	
20	1625-1649	61	47	108	
21	1650-1674	73	105	178	
22	1675-1699	13	24	37	
23	1700-1724	19	32	51	
24	1725-1749	46	47	93	
25	1750-1774	55	35	90	
26	1775-1799	11	30	41	
27	Eindtotaal	3363	964	4327	

regio\_periode wwe\_perid

Lijst met draaitabelvelden

Velden kiezen om toe te voegen aan rapport:

- ☒ id
- ☐ stad
- ☐ datum
- ☐ nr
- ☐ annotatie\_id
- ☐ pv\_type
- ☐ regio
- ☒ periodes\_25jaar
- ☒ wwe

Velden slepen tussen onderstaande gebieden:

Rapportfilter Kolomlabels

Rijlabels Waarden

periodes\_25jaar Aantal van id

Bijwerken van indeling ... Bijwerken

## En verder...

Excel biedt erg veel mogelijkheden om de gegevens uit je draaitabel verder vorm te geven, te groeperen of in een grafiek om te zetten. In de volgende afbeelding zie je hoe de frequenties in bovenstaande tabel omgezet zijn in percentages per tijdsdoorsneden

The screenshot shows the Microsoft Excel interface with a PivotTable and the PivotTable Field List. The PivotTable is located in the range B3:E27. The PivotTable Field List is on the right side of the screen.

**PivotTable Data:**

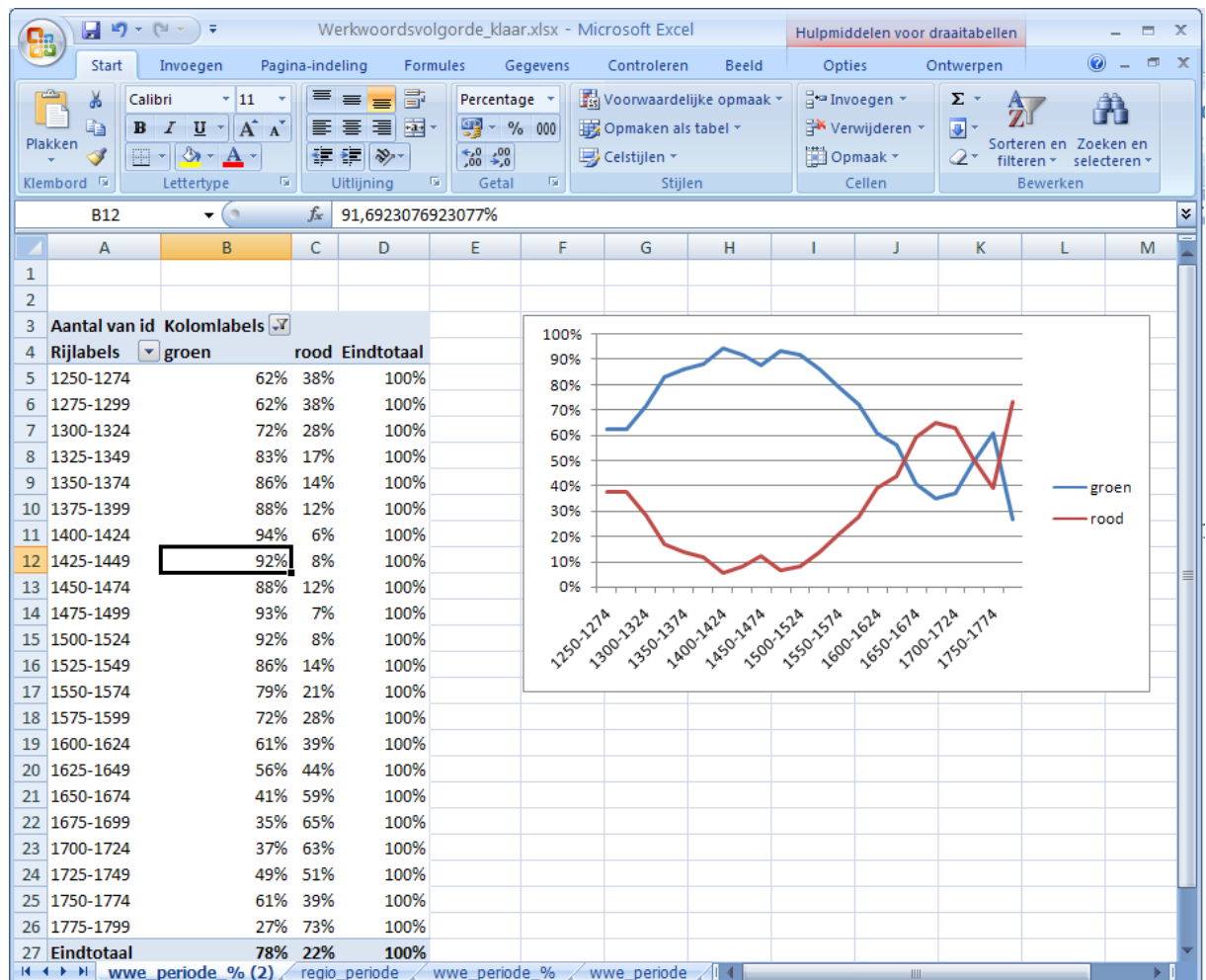
Rijlabels	groen	rood	Eindtotaal
1250-1274	62%	38%	100%
1275-1299	62%	38%	100%
1300-1324	72%	28%	100%
1325-1349	83%	17%	100%
1350-1374	86%	14%	100%
1375-1399	88%	12%	100%
1400-1424	94%	6%	100%
1425-1449	92%	8%	100%
1450-1474	88%	12%	100%
1475-1499	93%	7%	100%
1500-1524	92%	8%	100%
1525-1549	86%	14%	100%
1550-1574	79%	21%	100%
1575-1599	72%	28%	100%
1600-1624	61%	39%	100%
1625-1649	56%	44%	100%
1650-1674	41%	59%	100%
1675-1699	35%	65%	100%
1700-1724	37%	63%	100%
1725-1749	49%	51%	100%
1750-1774	61%	39%	100%
1775-1799	27%	73%	100%
<b>Eindtotaal</b>	<b>78%</b>	<b>22%</b>	<b>100%</b>

**PivotTable Field List:**

- Velden kiezen om toe te voegen aan rapport:
  - ☒ id
  - ☐ stad
  - ☐ datum
  - ☐ nr
  - ☐ annotatie\_id
  - ☐ pv\_type
  - ☐ regio
  - ☒ periodes\_25jaar
  - ☒ wwe
- Velden slepen tussen onderstaande gebieden:
  - Rapportfilter: wwe
  - Rijlabels: periodes\_25jaar
  - Waarden: Aantal van id

Zinvol zou ook zijn om die percentages in een grafiek af te beelden zodat je een diachrone trend in de werkwoordsvolgorde ziet verschijnen.





Uiteraard hangt de manier waarop je de data precies wil samenvatten nauw samen met de onderzoeksvragen die je jezelf hebt gesteld.

## Terminologie

Voor **algemene en taalspecifieke terminologie** verwijzen we o.a. naar de volgende werken:

Dubois, J., et alii (1994). *Dictionnaire de linguistique et des sciences du langage*. Paris: Larousse.

Finch, G. (2000). *Linguistic Terms and Concepts*. New York: St. Martin's Press.

Glossary of linguistic terms. <http://www.sil.org/linguistics/GlossaryOfLinguisticTerms/>

Hoek, T., J. Houtman & J. Jullens (1988). *Grammaticaal woordenboek*. Nijhoff.

Matthews, P. H. (1997). *The Concise Oxford Dictionary of Linguistics*. Oxford: Oxford University Press.

Trask, R. L. & P. Stockwell (2007). *Language and linguistics : the key concepts*. London; New York: Routledge.

Vrisekoop, M. C. (2001). *Grammaticale termen*. Coutinho.

Wortschatz Universität Berlin. [http://corpora.uni-leipzig.de/?dict=fra\\_mixed](http://corpora.uni-leipzig.de/?dict=fra_mixed)

Bij het lezen van anderstalige werken is het soms niet duidelijk of de terminologie tussen twee of meerdere talen helemaal overeenkomt. Voor **vergelijkende terminologie** tussen het Engels en Frans kan er verwezen worden naar de French/English Glossary of Linguistic Terms ([http://www.sil.org/linguistics/glossary\\_fe/glossary\\_index.asp](http://www.sil.org/linguistics/glossary_fe/glossary_index.asp)). Voor de andere talen geven we alvast enkele basisnoties mee (cf. Bijlage B).

## Bibliografie

- Anderson, W. J. & J. Corbett (2009). *Exploring English with online corpora : an introduction*. Basingstoke England ; New York: Palgrave Macmillan.
- Baker, P., A. Hardie & T. McEnery (2006). *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Habert, B. (2005). *Instruments et ressources électroniques pour le français*. Gap-Paris: Ophrys.
- Haspelmath, M. (2014). The Generic Style Rules for Linguistics.  
<http://www.eva.mpg.de/linguistics/staff/martin-haspelmath/generic-style-rules.html>  
(14 January, 2015)
- Johnson, K. (2008). *Quantitative methods in linguistics*. Malden (Mass.): Blackwell.
- Krug, M. G. & J. Schlüter (2013). *Research methods in language variation and change*. New York: Cambridge University Press.
- Litosseliti, L. (2010). *Research methods in linguistics*. New York (N.Y.): Continuum.
- Lüdeling, A., S. Evert & M. Baroni (2007). Using Web data for linguistic purposes. In Marianne, H., N. Nesselhauf & C. Biewer (eds.) *Language and Computers, Corpus Linguistics and the Web*. Rodopi. 7-24.
- McEnery, T., R. Xiao & Y. Tono (2006). *Corpus-Based Language Studies: An Advanced Resource Book*. New York: Routledge.
- Meyer, C. F., R. Grabowski, H.-Y. Han, K. Mantzouranis & S. Moses (2003). The World Wide Web as Linguistic Corpus. In Leistyna, P. & C. F. Meyer (eds.) *Language and Computers, Corpus Analysis: Language Structure and Language Use*. Amsterdam-New York: Rodopi. 241-54.
- Norde, M. (2007). Citeren en verwijzen in taalkundige opstellen: Het Harvardsysteem.  
<http://murielnorde.com/Harvardsysteem.pdf> (14 January, 2015)
- Paltridge, B. & A. Phakiti (2010). *Continuum companion to research methods in applied linguistics*. New York, N.Y.: Continuum.
- Rasinger, S. M. (2009). *Quantitative research in linguistics : an introduction*. London: Continuum.
- Renouf, A., A. Kehoe & J. Banerjee (2007). WebCorp: An Integrated System for Web Text Search. In Hundt, M., N. Nesselhauf & C. Biewer (eds.) *Corpus Linguistics and the Web*. Amsterdam - New York: Rodopi. 47-67.
- Sanders, E. (2011). Eerste Hulp Bij e-Onderzoek. <http://www.kb.nl/hpd/diensten/sanders-eerste-hulp-bij-e-onderzoek.pdf> (13 June, 2012)
- Spanoghe, A.-M. Vade-mecum pour un mémoire de licence en linguistique romane. La rédaction d'un texte scientifique (en collaboration avec E. Roegiest), Gent, Academia Press, 1997, 50 p.
- Teubert, W. & A. Cermáková (2007). *Corpus linguistics : a short introduction*. London: Continuum.
- Vrisekoop, M. C. (2001). *Grammaticale termen*. Coutinho.



## Bijlagen

### 3.1 Bijlage A: Criteria voor de Bachelor-/ Masterproef

1. Vertrekt de verhandeling van een duidelijke *status quaestionis* en/of probleemstelling?
2. Wordt de keuze van het onderwerp duidelijk geformuleerd?
3. Worden in de verhandeling de juiste methodologische stappen gezet om de gestelde problemen op te lossen?
4. Wordt de methodologie/ het theoretisch-methodologisch kader voldoende verantwoord?
5. Is de verhandeling duidelijk gestructureerd?
6. Geeft de verhandeling uiteindelijk een antwoord op de initiële onderzoeksvragen?
7. Is de uiteenzetting helder en coherent?
8. Worden de begrippen op een verantwoorde wijze gebruikt, d.w.z. kritisch geselecteerd, duidelijk gedefinieerd en consequent gehanteerd?
9. Worden de tussentijdse en eindconclusies voldoende ondersteund door het gepresenteerde materiaal, de analyses en de argumentatie?
10. Is de verhandeling een originele en/of relevante bijdrage tot het onderzoeksgebied?
11. Is de relevante literatuur op een adequate wijze verwerkt?
12. Is de omgang met het basismateriaal (teksten, corpora, data, bronnen allerhande) wetenschappelijk verantwoord?
13. In het geval van een teksteditie, zijn de uitgegeven teksten voldoende verklaard en geïnterpreteerd?
14. In het geval van een teksteditie, is voldoende zorg besteed aan de transcriptie van de uitgegeven teksten?
15. Zijn het notenapparaat en de bibliografie volledig, transparant, consistent?
16. Is de materiële presentatie van de scriptie verzorgd? (lay-out, typografie, tekstcorrectie, ...)
17. Is het taalgebruik in de scriptie (grammaticaal) correct?
18. Is het taalgebruik stilistisch adequaat: zakelijk, helder, wetenschappelijk?

## 3.2 Bijlage B: Grammaticale terminologie<sup>1</sup>

<b>zin</b>	<b>Satz</b>	<b>sentence</b>	<b>proposition</b>	<b>frase</b>
bijzin	Nebensatz	subordinate clause	proposition subordonnée	oración subordinada
hoofdzin	Hauptsatz	main clause	proposition principale	oración principal
woord	Wort	word	mot	palabra
woordgroep	Wortgruppe / Fügung	phrase	syntagme	sintagma
<b>zinsdelen</b>	<b>Satzglieder</b>	<b>parts of the sentence/</b>	<b>parties du discours</b>	<b>partes de la oración</b>
onderwerp / subject	Subjekt / Satzgegenstand	subject	sujet	sujeto
gezegde	Prädikat / Satzaussage	predicate	prédicat	predicado
naamwoordelijk deel v/h gezegde	Prädikatsnomen / Prädikativ	nominal part / subject complement	partie nominale du prédicat	atributo
lijdend voorwerp/ (direct) object	Akkusativobjekt/ direktes Objekt	direct object	objet direct	objeto directo
indirect voorwerp indirect object meewerkend voorwerp	Dativobjekt / indirektes Objekt	indirect object	objet indirect	objeto indirecto
voorzetselvoorwerp	Präpositionalobjekt	prepositional object (PP) / predicator complement	complément prépositionnel	complemento preposicional
bijwoordelijke bepaling	Adverbialbestimmung	adverbial adjunct	complément adverbial	complemento circunstancial o adverbial
bijvoegelijke bepaling	Attribut/ Beifügung	attributive adjunct	complément déterminatif	complemento predicativo
antecedent congruentie / vorm-overeenkomst	Antezedent / Bezugswort Kongruenz	antecedent concord /agreement	antécédent accord	antecedente congruencia / concordancia
<b>woordsoorten</b>	<b>Wortarten</b>	<b>parts of speech (PoS)</b>	<b>catégorie grammaticale</b>	<b>típos de palabras</b>
werkwoord	Verb/ Zeilwort	verb	verbe	verbo
zelfstandig werkwoord	Vollverb	lexical verb	verbe plein	verbo predicativo
koppelwerkwoord	Kopula	copula / linking verb	verbe copule	verbo copulativo
hulpwerkwoord	Hilfsverb	auxiliary verb	verbe auxiliaire	verbo auxiliar
zelfstandig naamwoord	Substantiv	noun	substantif/ nom	sustantivo/ nombre
bijvoeglijk	Adjektiv	adjective	adjectif	adjetivo

<sup>1</sup> Naar Vrisekoop (2001), herwerkt en gecorrigeerd.

naamwoord				
telwoord	Numerale / Zahlwort	numeral	numéral	numerales
hoofdtelwoord	Kardinalzahl/ Grundzahl	cardinal number / quantifier	nombre cardinal	numeral cardinal
rangtelwoord	Ordinale / Ordinalzahl	ordinal number	nombre ordinal	numeral ordinal
voornaamwoord	Pronomen / Fürwort	pronoun	pronom	pronombre
persoonlijk	Personalpronomen /	personal	pronom	pronombre
voornaamwoord	persönliches Fürwort	pronoun	personnel	personal
bezittelijk	Possessiv	possessive	pronom	adjetivo y
voornaamwoord		pronoun	possessif	pronombre posesivo
aanwijzend	Demonstrativ	demonstrative	pronom	adjetivo y
voornaamwoord		pronoun	démonstratif	pronombre demostrativo
betrekkelijk	Relativ	relative	pronom relatif	pronombre
voornaamwoord		pronoun		relativo
vragend	Interrogativ	interrogative	pronom	pronombre
voornaamwoord		pronoun	interrogatif	interrogativo
onbepaald	Indefinitpronomen	indefinite	pronom indéfini	pronombre
voornaamwoord		pronoun		indefinido
bijwoord	Adverb / Umstandswort	adverb	adverbe	adverbio
lidwoord / artikel	Artikel / Geschlechtswort	article	article	artículo
bepaald lidwoord / definiet artikel	bestimmter Artikel	definite article	article défini	artículo determinado
onbepaald lidwoord / indefiniet artikel	unbestimmter Artikel	indefinite article	article indéfini	artículo indeterminado
voorzetsel/prepositie	Präposition / Verhältniswort	preposition	préposition	preposición
achterzetsel / postpositie	Postposition	postposition	postposition	posición final
voegwoord	Konjunktion/ Bindewort	conjunction	conjonction	conjunción
tussenwerpsel / interjectie	Interjektion / Ausrufewort	interjection	interjection	interjección
<b>werkwoordsvormen</b>	<b>Verbalformen</b>	<b>verbal forms</b>	<b>formes verbales</b>	<b>formas verbales</b>
stamtijden	Stammformen	tenses	temps primitifs	formas verbales fundamentales
stam	Stamm	stem	racine/ radical	raíz / radical
onvoltooid tegen- woordige tijd (OTT)	Präsens	present tense	présent	presente
onvoltooid verleden tijd (OVT)	Präteritum	imperfect tense	imparfait	imperfecto
voltooid tegenwoordige tijd (VTT)	Perfekt	present perfect	passé composé	pretérito perfecto
voltooid verleden tijd (VVT)	Plusquamperfekt	past perfect	plus-que-parfait	pluscuamperfecto

persoonsvorm	Personalform	finite form	forme conjuguée/ finie	forma conjugada
onbepaalde wijs / infinitief	Infinitiv	infinitive	infinitif/ forme non-finie	infinitivo
tegenwoordig deelwoord	Partizip I	present participle	participe présent	gerundio
voltooid deelwoord	Partizip II	past participle	participe passé	participio
<b>wijs</b>	<b>Modus</b>	<b>mood</b>	<b>mode</b>	<b>modo</b>
aantonende wijs / indicatief	Indikatif	indicative	indicatif	indicativo
aanvoegende wijs / conjunctief	Konjunktiv	subjunctive	subjonctif	subjuntivo
gebiedende wijs / imperatief	Imperativ / Befehlsform	imperative	impératif	imperativo
naamval	Kasus/Fall	case	cas	caso
nominatief / onderwerpsvorm / subjectsform	Nominativ/ 1. Fall	nominative	nominatif	nominativo
genitief	Genitiv/ 2. Fall	genitive	génitif	genitivo
datief	Dativ/ 3. Fall	dative	datif	dativo
accusatief /voorwerpsvorm /objectvorm	Akkusativ/ 4. Fall	accusative	accusatif	acusativo
ablatief	Ablativ	ablative	ablatif	ablativo
vocatief	Vokativ/ Anredefall	vocative	vocatif	vocativo
<b>overige</b>				
actieve vorm/ bedrijvende vorm	Aktivform	active voice	voix active	activo
passieve vorm / lijdende vorm	Passivform / Leideform	passive voice	voix passive	pas ivo
trappen van vergelijking	Steigerungsstufen	degrees of comparison	degrés de comparaison	grado de los adjetivos calificativos
stellende trap	Positiv/ Grundstufe	positive	positif	el (grado) positivo
vergroten trap!	Komparativ/ Vergleichsstufe	comparative	comparatif	comparativo
overtreffende trap	Superlativ/ Höchststufe	superlative	superlatif	superlativo
toevoegsel/ affix	Affix	affix	affixe	afijo
voorvoegsel/ prefix	Präfix/ Vorsilbe	prefix	préfixe	prefijo
achtervoegsel/suffix	Suffix/ Nachsilbe	suffix	suffixe	sufijo
klinker/vocaal	Vokal/ Selbstlaut	vowel	voyelle	vocal
medeklinker / consonant	Konsonant / Mitlaut	consonant	consonne	consonante
lettergreep/ syllabe	Silbe	syllable	syllabe	sílaba